SOA Research Institute
SOCIETAL PURPOSE

Actuarial Innovation and Technology

From Health Inequities to Societal Bias: Insights from an Expert Panel on AI and Actuarial Responsibility

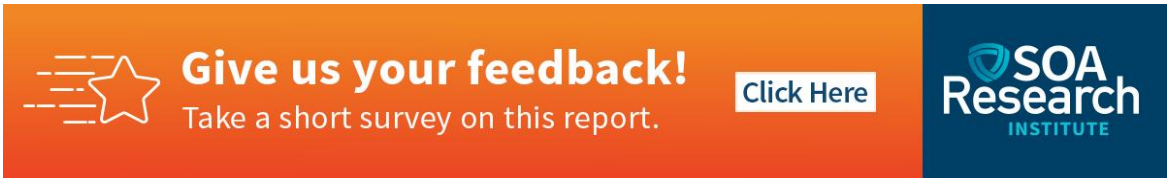
OCTOBER 2025

From Health Inequities to Societal Bias:

Insights from an Expert Panel on AI and Actuarial Responsibility

AUTHOR Ronald L Poon Affat, FSA, FIA, MAAA

SPONSOR Actuarial Innovation and Technology Strategic Research Program



SUMMARY

From Health Inequities to Societal Bias

Insights from an Expert Panel on AI and Actuarial Responsibility

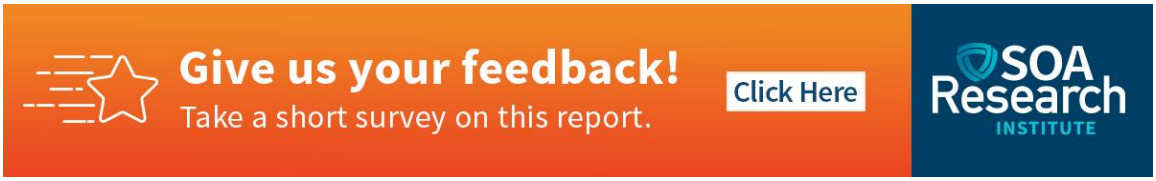
Executive Summary

The Society of Actuaries Research Institute convened a panel of experts to explore how potential bias in artificial intelligence is affecting the insurance industry—and how actuaries can evolve to meet this challenge. The discussion made clear that AI can reflect bias in data. Or as one panelist put it: "the data is not the data." In other words, what appears to be a neutral fact may carry the context, nuance, and social and behavioral science dimensions that shape its interpretation. Drawing from healthcare, regulatory, technical, and actuarial domains, the panelists explored how well-meaning models may fail when they treat historical data as objective truth. In high-stakes contexts like insurance and healthcare, these patterns may contribute to unwanted disparities if not addressed through improved practices. For actuaries, this is about more than technical rigor—it is about shaping ethical frontiers for the profession, ensuring that models designed or validated by actuaries promote fairness, trust, and accountability in an AI-driven future.

Key themes included:

- Historical Data and Objective Truth Are Different
  Legacy data reflects legacy decisions—not necessarily fairness. Treating it as neutral sustains any embedded past discrimination. Reassessing how data is sourced, contextualized, and interpreted can help avoid perpetuating potential past discrimination.
- Flawed Inputs Create Compounded Bias
  Using clinical assumptions like race-based kidney function scores or underwriting proxies like ZIP codes may encode real-world disparities into the modeling process. If left unexamined, AI may amplify these effects and scale them forward.
- Build Ethics Into Operations
  Articulating high-level principles like fairness and transparency are insufficient on their own. To be effective, principles of fairness and transparency must be embedded into daily decisions, model governance, and product design.
- Actuaries Need to Be More Like Social Scientists
  Traditional actuarial training emphasizes technical precision. But identifying possible sources of bias requires softer skills: understanding how social forces shape data, asking who benefits, and seeing beyond the surface-level patterns the data presents.
- Humans in the Loop
  When applied thoughtfully, AI can reduce diagnostic delays, expand access, and support underserved populations. These examples show what is possible when humans stay in the loop.

This report distills the panel's insights and highlights opportunities for actuaries and insurers to go beyond compliance—toward leadership in building AI systems that are equitable, accountable, and worthy of public trust.



Section 1 The Data Isn't Neutral: Rethinking AI Assumptions

1.1 QUESTIONING NEUTRAL DATA

The panel began with a multidisciplinary discussion that re-examined foundational assumptions in data science, actuarial practice, and sociotechnical system design—centered on how historical data, if treated as objective truth, can sustain potential systemic inequities in AI and insurance.

1.2 STRUCTURAL IMBALANCES IN DATA

At the outset, the panel noted that the notion of historical data as neutral or objective is problematic. The panel asserted that historical datasets may not represent the full spectrum of experiences in society. Rather, such data may disproportionately reflect the realities and perspectives of dominant or historically empowered groups. In this framing, data does not merely describe the world; it reflects any structural imbalances of the past. Accordingly, when AI systems are trained on such biased datasets, the outputs can carry forward potential systemic exclusions if left unaddressed.

1.3 LIMITS OF BIG DATA

One participant highlighted the fallacy of assuming that simply building systems on "big" or "validated" historical data would yield fair results. If the input data fails to account for the lived realities of potentially underrepresented or underserved groups, the resulting AI systems will perpetuate those discrepancies or inequities. The concern was not limited to training data alone but extended to the very benchmarks and metrics used to evaluate model performance. AI validation processes that rely on the same historical datasets risk reinforcing existing disparities rather than correcting them.

1.4 EQUITY VS. EFFICIENCY

The panel explored how efficiency-oriented design incentives in AI compound this problem. Systems are often built to scale and prioritize performance, but without a parallel consideration of for whom they perform. If data is skewed toward groups that have historically benefited from system design, optimization may come at the expense of marginalized populations. To mitigate risks of perpetuating potential inequities, the panel recommended a shift in design philosophy: rather than retrofit equity after performance, build it in from the start.

1.5 BIASED PROXIES

A concrete illustration arose in the critique of data proxies that have been shown to replicate discriminatory structures. One example discussed in detail was the use of ZIP codes. While ZIP codes may seem innocuous, they often correlate strongly with race, income, and historical redlining.[1] Education level, credit scores, and education were also cited as common proxies for protected class characteristics.[2] Although these variables are technically "non-sensitive," their inclusion in AI models may inadvertently replicate patterns associated with discrimination. To

[1] Federal Trade Commission, *Big Data: A Tool for Inclusion of Exclusion? Understanding the Issues* (Washington, DC: Federal Trade Commissions, January 2016), https://www.ftc.gov/reports/big-data-tool-inclusion-or-exclusion-understanding-issues-ftc-report.
[2] National Association of Insurance Commissioners. *NAIC Model Bulletin: Use of Artificial Intelligence Systems by Insurers.* Draft, December 2, 2023. Adopted by Executive (EX) Committee and Plenary, December 4, 2023; adopted by the Innovation, Cybersecurity, and Technology (H) Committee, December 1, 2023, https://content.naic.org/sites/default/files/cmte-h-big-data-artificial-intelligence-wg-ai-model-bulletin.pdf.pdf.

limit the potential for replicating such patterns, the panel suggested carefully examining such relationships during both model development and deployment.

1.6 BROADENING THE TOOLKIT

The panel suggested that professionals in actuarial science and insurance expand their toolkit beyond quantitative techniques. Social science, behavioral economics, and even historical analyses were suggested as useful complements to traditional statistical modeling. The aim is not to replace technical expertise, but to supplement it with broader insights that may help anticipate and reduce the risk of unintended effects. For instance, they could consider themselves as "social scientists" working within—and influencing—social systems.

1.7 CULTURAL AND LINGUISTIC BIAS

One part of the discussion focused on how seemingly neutral data points may carry cultural or racial associations. For example, a participant referenced the significant volume of ongoing research into how different dialects are interpreted in medical contexts. Phrases like "uncooperative" or "belligerent" in electronic health records can reflect clinician bias—especially when patients speak in African American Vernacular English or have distinct speech patterns. These notes, if used as features in predictive machine learning models, could unintentionally shape outcomes. Similarly, concerns were raised about emerging models that use voice inflection to predict health outcomes. The panel questioned the cultural assumptions behind such inputs and suggested careful evaluation before incorporating them into predictive systems.

1.8 ALGORITHMS REFLECT SOCIETY

In referencing external work, the panel cited *Algorithms of Oppression* by Safiya Noble to illustrate how algorithmic outputs can reflect and amplify societal biases. One example discussed from the book involved how search engine results for terms like "Black girls" historically yielded inappropriate or stereotyped content when autocompleting the phrase.[3] The key takeaway was that algorithms are not inherently neutral—they mirror the values, assumptions, and blind spots of their designers rather than objective truths. Recognizing this possibility can help guide more thoughtful system development.

1.9 KEY TAKEAWAYS

- Historical data may not be an impartial record but a partial one—it may encode the experiences of those who held power, overlooking or misrepresenting marginalized communities. Avoiding the building of AI systems that unintentionally replicate those same exclusions requires conscious correction.
- Systems designed primarily for performance and scale may not work equally well across all populations. Reducing the potential for overlooking underrepresented groups requires introducing equity considerations early in system design.
- Complementing the technical rigor of actuarial science with a critical understanding of the social context in which data is created and used may provide useful context for interpreting patterns in the data that may lie below the surface, anticipating where results could have broader implications, and reducing unintended effects.

[3] Noble, Safiya Umoja, *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018. https://nyupress.org/9781479837243/algorithms-of-oppression/

Section 2 Diagnosis: Biased Data, Unequal Outcomes

2.1 BIAS IN HEALTHCARE DATA

The panel discussion on algorithmic bias in healthcare focused on how data-driven tools influence decisions about patient care and insurance coverage. Rather than assuming healthcare data is inherently objective, the panel examined how existing datasets may reflect historical inequalities in treatment and access. They pointed out that healthcare, while often seen as data-rich and evidence-based, is not exempt from the social and systemic influences that shape how care is delivered and recorded—factors that can affect the performance and fairness of AI systems trained on such information

2.2 POTENTIAL SYSTEMIC INEQUITIES IN CLINICAL RECORDS

The panel began by explaining that healthcare is frequently misunderstood as a domain where neutrality can be assumed simply because decisions are rooted in data such as lab results, patient histories, or clinical records. However, this presumption overlooks how the data itself is shaped by any potential systemic inequalities in access, treatment, and outcomes. Data, the panel pointed out, does not emerge in a vacuum. It is shaped by the real-world conditions under which care is delivered. Consequently, AI systems trained on such data are liable to reflect those inequities unless deliberate steps are taken to address them.

2.3 CASE STUDY: CARE MANAGEMENT ALGORITHM

A real-world case helped illustrate this dynamic. The panel discussed a 2019 healthcare algorithm designed to determine which patients should be prioritized for additional care management. At first glance, the model appeared data-driven and fair—it used past healthcare spending as a proxy for future health needs. However, this design choice had significant implications: because Black patients have historically received less care and incurred lower medical costs, the algorithm under-flagged them for additional services. Despite equal or greater clinical needs, these patients were not identified by the system as needing extra support.[4] This example demonstrated how a seemingly neutral input—past spending—can embed bias that excludes the very populations that should be prioritized.

2.4 CASE STUDY: DERMATOLOGY AI

Another example focused on dermatological AI tools. The panel explained that a diagnostic model had been trained primarily on images of white skin. As a result, it was significantly less effective at identifying skin conditions on people with darker skin tones.[5] The failure was not malicious in intent, but rather the product of narrow data sourcing and oversight in evaluating how model performance would generalize across diverse populations. This highlighted the risk of overlooking cultural, racial, or demographic diversity when curating datasets or evaluating algorithmic outputs.

2.5 ACTUARIES' EXPANDING ROLE

The panel broadened the discussion to actuaries specifically, drawing attention to their growing role in shaping healthcare outcomes—not only through traditional pricing models, but also via the design and evaluation of

---

4 Ziad Obermeyer, Brian Powers, Christine Vogeli, and Sendhil Mullainathan, "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations," *Science* 366, no. 6464 (October 25, 2019): 447–453, https://www.science.org/doi/10.1126/science.aax2342.
5 Roxana Daneshjou et al., "Disparities in Dermatology AI Performance on a Diverse, Curated Clinical Image Set (DDI Dataset)," *arXiv* (March 2022), https://arxiv.org/abs/2203.08807.

predictive models that influence access to services. In contexts such as health insurance, algorithms are now used to determine who is recommended for further diagnostic testing, who qualifies for a specialist referral, or which procedures are reimbursed. The panel noted that while actuaries may not always design these algorithms, they are often in decision-making or decision-influencing roles where they review model outputs, integrate them into insurance frameworks, or advise on their implementation. As such, they need to be mindful of the assumptions behind model design—especially the choice of proxies and benchmarks that might unintentionally reinforce potential disparities.

2.6 BIAS FROM OMISSION, NOT INTENT

The panel also addressed the misconception that bias always results from intentional discrimination. Instead, many of the examples discussed—whether based on healthcare spending or skin tone—resulted from technical decisions made without full awareness of their implications. In such cases, the harm comes not from malice, but from omission: the failure to ask whose data is being used, whose experience is missing, and whether the model will serve all populations equally well.

2.7 PROGRESS AND EVOLVING RESPONSIBILITIES

Encouragingly, the panel pointed to areas where progress is possible and already underway. There was acknowledgment that the industry is becoming more attuned to these issues, and that asking the right questions—early and often—can shape outcomes in meaningful ways. The discussion recognized that actuaries, previously viewed as technical advisors, are now participating in strategic conversations about benefit design and model governance. This evolving role allows them to raise concerns, suggest improvements, and support efforts to embed fairness within interdisciplinary teams.

2.8 KEY TAKEAWAYS

- Healthcare data is not inherently neutral—it may reflect societal inequalities. The panel emphasized that clinical datasets are shaped by actual patterns in access, treatment, and outcomes, which may reflect disparities rather than objective truths. AI models trained on this data risk reproducing any inequities unless practitioners actively account for missing or skewed perspectives.
- Seemingly neutral inputs can encode bias. Real-world examples—like algorithms that under-prioritized Black patients due to lower historical healthcare spending—illustrated how flawed proxies can lead to unfair outcomes. These were not cases of malicious design, but of unexamined assumptions that ignored the social context behind the data.
- Actuaries have a growing responsibility in shaping AI outcomes. While they may not always design algorithms, actuaries increasingly review, integrate, and advise on predictive tools that affect healthcare access. The panel encouraged the profession to scrutinize proxies, advocate for disaggregated metrics, and participate in cross-functional efforts to ensure fairness.

Section 3 Race-Based Adjustments and Their Legacy in Medical Data

3.1 RACE-BASED CLINICAL TOOLS IN MEDICINE

The panel discussion around the question of how race-based clinical tools could have influenced healthcare disparities—and how those potentially embedded biases could transfer into AI training data—unfolded with extensive examples, technical insight, and a measured exploration of the harms and improvements emerging in the use of artificial intelligence within the medical field.

3.2 EGFR AND KIDNEY FUNCTION

The panel observed that race-based clinical adjustments are deeply rooted in historical inaccuracies that persist in contemporary medicine. The discussion focused on how certain clinical tools and calculators have incorporated race without valid scientific basis, leading to real-world consequences for care delivery and outcomes.[6]

A major example discussed was the estimated glomerular filtration rate (eGFR),[7] a measure used to estimate kidney function. The panel described how, in 1999, the Modification of Diet in Renal Disease (MDRD) study group published a new equation to calculate the eGFR. In the research study, Black participants had higher creatinine levels. Rather than investigate the causes or question the validity of this observation, the researchers introduced a correction factor—premised on the assumption that Black individuals have inherently higher muscle mass.[8] This notion, rooted in outdated and prejudiced beliefs, led to two different eGFR standards: one for Black individuals and one for non-Black individuals.

The panel delved deeper into the specific impacts of removing the race factor from the eGFR calculation. They described a detailed case: for a Black person with a creatinine level of 1.8, using the race-based MDRD equation yielded an eGFR of 53, whereas the race-neutral Chronic Kidney Disease Epidemiology Collaboration (CKD-EPI) calculation, adopted in 2021, gave a significantly lower value of 36. This discrepancy shifted the patient from one stage of chronic kidney disease to a more severe one, which in turn would drastically alter treatment decisions and urgency. The panel made clear that this kind of difference is not abstract—it determines whether someone receives timely intervention or faces progression to more advanced stages of kidney disease.

3.3 OTHER RACE-BASED GUIDELINES

The consequences of this divergence were profound. Because the adjusted formula overestimated kidney function in Black patients, many were shown as healthier than they actually were. As a result, they received delayed treatment, with some progressing to kidney failure before appropriate interventions could be made. While the panel noted that this race correction factor was officially removed in 2021, the inertia of healthcare systems means the change has taken time to make its way into full practice. This lag means that models trained on older datasets may still reflect and perpetuate that historical bias for years to come.[9]

---

[6] Darshali A. Vyas, Leo G. Eisenstein, and David S. Jones, "Hidden in Plain Sight — Reconsidering the Use of Race Correction in Clinical Algorithms," ed. Debra Malina, *New England Journal of Medicine* 383, no. 9 (2020): 874–82, https://doi.org/10.1056/nejmms2004740.

[7] Andrew S. Levey, John P. Bosch, J. Bryan Lewis, Tamara Greene, Neil Rogers, and Daniel Roth, "A More Accurate Method to Estimate Glomerular Filtration Rate from Serum Creatinine: A New Prediction Equation. Modification of Diet in Renal Disease Study Group," *Annals of Internal Medicine* 130, no. 6 (March 16, 1999): 461–70, doi:10.7326/0003-4819-130-6-199903160-00002.

[8] S. Ahmed, C. T. Nutt, N. D. Eneanya, et al., "Examining the Potential Impact of Race Multiplier Utilization in Estimated Glomerular Filtration Rate Calculation on African-American Care Outcomes," *Journal of General Internal Medicine* 36 (2020), https://doi.org/10.1007/s11606-020-06280-5

[9] Cynthia Delgado, Mallika Baweja, Deidra C. Crews, et al., "A Unifying Approach for GFR Estimation: Recommendations of the NKF-ASN Task Force on Reassessing the Inclusion of Race in Diagnosing Kidney Disease," *American Journal of Kidney Diseases* 79, no. 2 (2021), doi:https://doi.org/10.1053/j.ajkd.2021.08.003.

Importantly, the discussion emphasized that this is not an isolated case. Other guidelines used in medicine still rely on race-based inputs today. One such tool, the 2012 Global Lung Function Initiative (GLI), calculates lung function using different reference values for various ethnicities. Even though genetic research has shown that human organs are essentially the same across racial groups, medical guidelines still treat lung capacity differently depending on race.[10] In 2022, the race-neutral GLI Global was published[11] and in 2023, the American Thoracic Society recommended the use of the GLI Global reference values.[12]

CONSTRUCTIVE USE OF AI IN RADIOLOGY

However, amid the serious concerns raised, the panel also emphasized a powerful example of how AI is being used constructively in healthcare. In radiology, generative AI has been deployed to assist in reading medical images such as x-rays and computed tomography (CT) scans. A recent case study from Northwestern University was referenced, where AI tools helped radiologists identify life-threatening conditions more quickly.[13]This improvement enabled them to prioritize urgent cases while also improving overall throughput. The panel noted that radiologists using this system increased their case completion rate by over 15% without any loss of accuracy.[14]

3.4 IMPACT ON UNDERSERVED COMMUNITIES

This radiology application was seen as especially promising for underserved communities. In regions with a shortage of radiologists, the technology offers a way to expand access to faster, more responsive care. By triaging the most critical images, AI helps reduce delays in diagnosis, potentially saving lives. The panel welcomed this as a meaningful example of AI being deployed with clear benefits that appear to treat all populations equitably.

3.5 CORRECTING KIDNEY DONOR BIAS

Another example was raised concerning the evaluation of donor kidneys. A scoring system used to determine the quality of deceased donor kidneys historically assigned a negative weight to Black donors that predicted their organs would fail at a higher rate than non-Black donors.[15] Deaths due to heart disease, stroke, or diabetes were treated as less impactful on kidney quality than race alone. This approach reduced the number of viable kidneys available for transplantation to patients who desperately needed them. The panel reported that this race-based discounting has since been removed as of 2024.[16]

---

[10] Philip H. Quanjer, Sanja Stanojevic, Tim J. Cole, et al., "Multi-Ethnic Reference Values for Spirometry for the 3–95-yr Age Range: The Global Lung Function 2012 Equations," *European Respiratory Journal* 40, no. 6 (2012): 1324–43, doi:https://doi.org/10.1183/09031936.00080312.

[11] Charles Bowerman, Nirav R. Bhakta, Davide Brazzale, et al., "A Race-Neutral Approach to the Interpretation of Lung Function Measurements," *American Journal of Respiratory and Critical Care Medicine* 207, no. 6 (2023): 768–74, doi:https://doi.org/10.1164/rccm.202205-0963oc.

[12] American Thoracic Society, "ATS Publishes Official Statement on Race, Ethnicity and Pulmonary Function Test Interpretation," *American Thoracic Society*, accessed September 19, 2025, https://site.thoracic.org/about-us/news/ats-publishes-official-statement-on-race-ethnicity-and-pulmonary-function-test-interpretation.

[13] Jianhua Huang, Matthew T. Wittbrodt, Charles N. Teague, et al., "Efficiency and Quality of Generative AI–Assisted Radiograph Reporting," *JAMA Network Open* 8, no. 6 (2025): e2513921, doi:https://doi.org/10.1001/jamanetworkopen.2025.13921.

[14]Northwestern University, "New AI Transforms Radiology with Speed, Accuracy Never Seen Before," *Northwestern Now*, May 29, 2025, https://news.northwestern.edu/stories/2025/06/new-ai-transforms-radiology-with-speed-accuracy-never-seen-before.

[15] Jennifer Miller, Gregory R. Lyden, William T. McKinney, Jon J. Snyder, and Ajay K. Israni, "Impacts of Removing Race from the Calculation of the Kidney Donor Profile Index," *American Journal of Transplantation* 23, no. 5 (2023): 636–41, doi:https://doi.org/10.1016/j.ajt.2022.12.016.

[16] "OPTN Board Approves Exclusion of Race, Hepatitis C Status from Estimate of Deceased Donor Kidney Function," *OPTN*, Health Resources & Services Administration, published 2024, https://optn.transplant.hrsa.gov/news/optn-board-approves-exclusion-of-race-hepatitis-c-status-from-estimate-of-deceased-donor-kidney-function/.

3.6 KEY TAKEAWAYS

- Race-based clinical tools can distort treatment decisions. The panel examined how race corrections in calculators like eGFR and lung function reference values lacked solid scientific grounding and contributed to delayed or unequal care for Black patients. Even after their formal removal, these formulas continue to shape legacy datasets and influence AI models trained on outdated assumptions.
- Embedded bias in training data has long-term consequences. When medical algorithms rely on historical data that includes race-based adjustments that are no longer used, they risk replicating past disparities in new forms. The panel emphasized that without careful review of input variables and context, AI can carry forward—and even amplify—the consequences of flawed clinical assumptions.
- AI can also be part of the solution when applied thoughtfully. The panel cited radiology tools that help triage urgent cases and expand diagnostic capacity in underserved regions. These examples show how AI, when designed with equity in mind, can improve access and efficiency—especially where clinical resources are limited.

Section 4 Insurers, AI, and the Regulatory Challenge

4.1 PROGRESS ON AI BIAS IN INSURANCE REGULATION

When examining the insurance industry's progress in addressing AI bias through the lens of regulation, the panel highlighted encouraging signs of momentum—particularly in health insurance, where awareness and early reforms are starting to take shape. While challenges remain, there was recognition that both insurers and regulators are increasingly acknowledging the importance of fairness, transparency, and accountability in AI-driven decision-making.

4.2 BIAS, FAIRNESS, AND REGULATORY OVERSIGHT

The discussion opened by noting that bias in areas like healthcare can lead to uneven outcomes that affect access, treatment, and coverage. Panelists emphasized that addressing these issues is not only a matter of fairness but also falls squarely within the scope of regulatory oversight. While some insurers have started to incorporate these concerns into their practices, there remains a gap between recognizing the issue and implementing consistent, system-wide responses. Panelists observed that regulatory frameworks seem to be evolving alongside industry awareness.

4.3 DEFINING BIAS: HARMFUL VS. ACCEPTABLE

A central theme was that many insurers still conflate the term "bias" with something inherently negative, without distinguishing between harmful and benign (or even helpful) biases. From a regulatory standpoint, this lack of distinction leads to inconsistent interpretations. Laws and guidance often aim to prevent unfair discrimination, but if industry actors are unclear about what qualifies as problematic bias versus acceptable heuristics, their compliance efforts may end up being overly cautious in some areas and inattentive in others.

4.4 RETHINKING WHAT "PREDICTIVE" MEANS

The panel also questioned the term "predictive model," noting that many such models primarily reflect historical data and established patterns rather than making genuine forecasts. This distinction has regulatory significance, as it calls into question claims that model outputs are inherently objective or future oriented. If oversight processes rely too heavily on the label "predictive," they may overlook potential embedded historical biases that can shape a model's results. Focusing on how models perform in practice, rather than relying on industry terminology that may obscure underlying limitations, may mitigate this risk.

4.5 CHALLENGES IN APPLYING FAIRNESS AND ACCOUNTABILITY

The panel noted that insurers continue to face challenges in applying principles like fairness, transparency, and accountability in a consistent and operationally meaningful way. Part of the difficulty stems from ambiguous or inconsistently applied definitions. For example, "algorithmic bias" is often mistakenly attributed to the algorithm itself, rather than to the underlying data, assumptions, or human decisions that shape it. This lack of clarity can contribute to uneven practices across the industry.

4.6 THIRD-PARTY VENDORS AND COMPLIANCE RESPONSIBILITY

The discussion included the role of third-party vendors, which are playing an increasingly important part in providing data and modeling tools to insurers. Currently, insurers determine their standards for vendors and conduct vendors of audits accordingly. Panelists observed that insurers generally prioritize operational efficiency, while regulators

generally prioritize safeguarding consumer interests. Developing industry-wide vendor standards that reconcile these priorities would help clarify compliance responsibilities.

4.7 ASOPS IN THE CONTEXT OF COMPLIANCE

The panel noted that some insurers reference actuarial standards of practice (ASOPs) when responding to regulatory inquiries. In certain cases, this has included the suggestion that adhering to ASOPs satisfies broader legal obligations related to fairness. The panel pointed out that this interpretation may overlook the fact that ASOPs defer to legal requirements in the event of conflict. Clearer understandings of the respective roles of professional standards and regulatory law could support more consistent compliance practices.

4.8 BIAS WITHOUT INTENT

The panel remarked that these disparities do not require malice to be harmful. Models and practices that appear neutral can perpetuate potential systemic inequalities, and correcting those inequalities is not itself discriminatory. Failing to examine such root causes may inadvertently preserve the very inequities that are intended to be avoided.

4.9 BEYOND ALGORITHMS: HUMAN BEHAVIOR AND CULTURAL CHANGE

In closing, the panel offered a reflective perspective: the best long-term solution may not lie solely in refining algorithms, but in changing human behavior. If society can reduce the negative biases that humans bring into decision-making, this will eventually improve the data used in insurance models.

4.10 KEY TAKEWAYS

- Growing awareness but uneven application
  The panel observed that insurers and regulators are increasingly acknowledging the importance of fairness and transparency in AI. Some firms have made early progress, but translating abstract principles into operational practice still proves challenging.
- Definitions, labels, and accountability gaps
  Ambiguities in terms like “bias” and “predictive model” continue to complicate compliance efforts. Focusing on how models perform in practice rather than relying on industry framing and extending accountability the algorithms to data sources, assumptions, and third-party vendors may help to improve compliance.
- The need for cultural as well as technical change
  Addressing potential bias requires more than algorithmic tuning—it calls for deeper institutional awareness and behavioral change by the human decision-makers who shape and deploy them.

Section 5 The Actuary's Ethical Frontier

5.1 SHAPING ETHICAL FRONTIERS

In response to the question regarding how actuaries can navigate emerging ethical challenges in the era of AI—particularly moving beyond ASOP compliance to play a more active role in risk mitigation and public trust—the panel offered a well-rounded discussion. The exchange emphasized that actuarial professionals are uniquely positioned to influence how AI is adopted in insurance, not merely through technical expertise but through ethical leadership, critical thinking, and cross-disciplinary engagement.

5.2 BEYOND MINIMUM STANDARDS

The panel began by discussing common misunderstandings about the role of ASOPs. While these standards outline essential requirements for professional conduct, the panel described them as a baseline rather than the upper limit of ethical responsibility. They view compliance as a starting point, not the entirety of their duty. In rapidly evolving areas such as AI—where technology advances faster than formal guidance can keep pace—depending only on codified rules may leave important issues unaddressed. The panel highlighted that many ethical considerations related to algorithmic bias involve judgment calls that extend beyond what technical documentation or compliance checks can capture.

5.3 STEWARDSHIP AND PROFESSIONAL INTEGRITY

The panel views involvement with AI systems as an opportunity for actuaries to practice their role as professional stewards of judgment and integrity. The discussion highlighted that actuarial work is not limited to running models and reviewing technical output; it also involves exercising ethical discretion and assessing broader societal impacts. Actuaries can ask probing questions when they observe unintended or inequitable outcomes—even when a model technically meets validation criteria. Raising concerns when model outputs appear biased or when data sources are incomplete or skewed was described as part of the actuary's professional duty.

5.4 EXPLAINABILITY AS OVERSIGHT

The panel then explored the importance of transparency and explainability in AI-driven decisions. In many insurance environments, there is a risk that algorithmic models become opaque, particularly when outsourced or embedded in third-party systems. Retaining visibility into these systems is important for understanding outputs. Merely relying on performance metrics or vendor assurances is insufficient when the decisions generated by these tools affect real people—especially when those decisions involve coverage access, pricing, or claims adjudication.

5.5 CROSSING DISCIPLINARY BOUNDARIES

A key idea in the panel's exchange was the desire for actuaries to work across disciplines. It was acknowledged that the most consequential aspects of AI—including fairness, accountability, and trust—cannot be resolved through statistical calibration alone. These issues straddle ethics, social justice, behavioral science, and regulatory policy. The panel emphasized that actuaries must step outside the actuarial silo and engage with professionals in other fields who understand how various forms of bias or discrimination operate in practice and how vulnerable communities experience the results. This kind of interdisciplinary collaboration was framed not as a detour from actuarial work but as a necessary evolution of it.

5.6 HUMAN JUDGMENT AT THE CORE

To prevent AI from becoming a shield for impersonal decision-making, the panel highlighted the importance of keeping the "actuary in the loop." This goes beyond maintaining empathy—it includes the professional judgment and courage to question how modeling choices such as data selection, proxy variables, and objective functions embed human assumptions with ethical implications. The discussion stressed that behind every model are human hands, and when actuaries remain in the loop, those hands can be guided by professional values as well as technical objectives.

5.7 KEY TAKEAWAYS

- Ethical responsibility beyond ASOP compliance—The panel underlined that ASOPs provide a baseline, not the full extent, of an actuary's role in the AI era. Compliance alone may leave important ethical issues unaddressed, particularly as technology evolves faster than formal guidance. The panel believes that actuaries must be prepared to apply professional judgment to areas that lie outside codified rules, especially when algorithmic bias and fairness are at stake.
- Explainability and oversight in AI systems—A clear understanding of how AI models make decisions, whether developed in-house or by third parties, is important to understanding its output. Explainability is essential to ensure that coverage, pricing, and claims decisions can be justified and defended. Relying solely on vendor assurances or performance metrics may not be sufficient.
- Keeping the actuary in the loop—The panel emphasized that effective actuarial involvement goes beyond technical validation to include questioning data choices, proxy variables, and objective functions that embed human assumptions. Retaining the "actuary in the loop" helps ensure that models are guided by professional values as well as statistical goals. To achieve fair, accountable, and trustworthy AI models, cross-disciplinary collaboration—with experts in ethics, social science, and regulation—was framed as a necessary evolution of the profession.

Section 6 Conclusions

6.1 ETHICS IN EVERYDAY PRACTICE

The closing discussion tied earlier ethical and regulatory themes to the day-to-day realities of insurance, healthcare, and financial services. The panel stressed that bias, fairness, and access issues cut across geographies, sectors, and product lines—whether in healthcare, personal protection, asset insurance, or access to capital. These industries share a core mission: enhancing financial security and societal well-being, which depends on constant vigilance to spot where inequities may arise, from underwriting to claims, often at pivotal moments in people's lives.

6.2 THE DATA BEHIND THE DATA

A central refrain was "the data is not the data." Numbers in technical models are shaped by human actions, institutional practices, and social systems. Bias enters through the ways information is generated and recorded, not spontaneously. Addressing this requires looking beyond mathematical precision to understand the social conditions that produce the data. Actuaries, often trained to value precision to many decimal places, are well-positioned to ask why the data looks the way it does and whose experiences are absent or distorted. Embedding sociological and behavioral insights into actuarial work can surface these drivers before they become embedded in automated systems.

6.3 CROSS-INDUSTRY FAIRNESS

The panel noted that fairness problems are cross-industry, so solutions must be too. Incorporating user-specific context—through industry standards, additional research, or direct consumer input—can make outputs more relevant and reduce blind spots from relying solely on historical patterns. Clinical data was cited as an example: even with active reforms, many datasets still carry inequities from past collection practices and awareness of these legacies is essential for both model builders and regulators.

Section 7 Acknowledgments

The researchers' deepest gratitude goes to those without whose efforts this project could not have come to fruition: the Expert Panel participants and others for their diligent work overseeing questionnaire development, analyzing, and discussing respondent answers, and reviewing and editing this report for accuracy and relevance.

Dorothy Andrews, PhD, ASA, MAAA, CSPA– Senior Behavioral Data Scientist and Actuary at the NAIC, supporting the work of its Big Data AI Working Group.

Miranda Bogen, MA – Director at the Center for Democracy & Technology; expert in algorithmic accountability, AI bias, and civil rights impacts of technology.

Jurnell Cockhren – President & Founder of Civic Hacker LLC; expert in ethical AI, software engineering, and data justice.

Dale Hall, FSA, MAAA, CFA, CERA – Managing Director of Research at the Society of Actuaries; specializes in insurance research, actuarial science, and public policy.

Kimberly Sapre, DMSc, PA-C – Clinical manager; expert in healthcare data equity, digital health, and ethical implementation of AI in clinical environments.

Ronald Poon Affat, FSA, FIA, MAAA, CFA – Independent Board Director and Cross Continental Actuary; expert in AI governance, reinsurance, and emerging market insurance.

At the Society of Actuaries Research Institute:

Korrel Crawford, Senior Research Administrator

R. Dale Hall, FSA, MAAA, CFA, Managing Director of Research

Appendix A: Expert Panel Discussion Questions

The panel was asked the following questions to prompt discussion.

- Your career has expanded into neuroscience labs, code repositories, and civic movements. And you've seen firsthand that bias just isn't in the data. It's embedded in the systems' assumptions, incentives behind our technologies. In your view, what are the most common misconceptions about treating historical data as objective truth in AI development? And how should the insurance industry rethink their role to ensure AI advances equity?

- Your work in AI governance has shown that bias isn't just theoretical. It has real consequences. And of course, nowhere is this more visible and impactful than healthcare, where there's overwhelming evidence that racial and socioeconomic disparities persist in outcomes, access, and treatment quality. So why is healthcare such an urgent example of how algorithmic bias can reinforce inequality? And what should actuaries—whose models influence health insurance decisions—understand about how these disparities persist even when using data-driven systems that may appear neutral on the surface?

- Can you share how the medical field's use of race-based adjustments or other flawed clinical assumptions have contributed to disparities in care, and how those biases—once embedded in practice—carry over into the data used to train AI models? And conversely, is there an example where AI has been applied thoughtfully to improve diagnosis, treatment, or health equity for underserved populations?

- As an actuary and someone who has consistently emphasized distinguishing between harmful and helpful bias in AI: What stands out to you as the most meaningful progress insurers have made in addressing harmful bias? And where do you see the industry still falling short—particularly in turning principles like fairness, transparency, and accountability into consistent operational practice? Why do you think many insurers still underestimate or overlook the real-world risks and complexity of bias in AI systems?

- Actuaries have long served as stewards of risk and protective solvency, but as AI becomes more deeply embedded in insurance decision-making, the profession faces a new set of ethical challenges. Many concerns surrounding algorithmic bias are subjective, evolving, and not easily addressed by technical standards alone. How can actuaries position themselves to go beyond simply meeting ASOP requirements and take a more active role in identifying and mitigating risks? And how can they bring the human touch needed to ensure that AI-driven practices promote fairness and strengthen societal trust?

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
8770 W Bryn Mawr Ave, Suite 1000
Chicago, IL 60631
www.SOA.org