Lemonade

Lemonade Insurance Company   William D. Latza
5 Crosby Street, 3rd Floor   General Counsel and Secretary
New York, NY 10013

July 21, 2020   Via email to
                                                              dmatthews@naic.org,

Hon. Jon Godfread, Chair
Hon. Elizabeth Kelleher Dwyer, Vice Chair
Innovation and Technology (EX) Task Force
National Association of Insurance Commissioners
1100 Walnut Street, Suite 1500
Kansas City, MO 64106

   Re: NAIC Principles on Artificial Intelligence

Dear Commissioner Godfread and Superintendent Dwyer:

Thank you for the opportunity to comment on the Principles on Artificial Intelligence as adopted by the Artificial Intelligence (EX) Working Group at its June 30, 2020 teleconference meeting. We applaud the NAIC for undertaking this difficult task, and for the open process that has brought us to this point so quickly.

Lemonade is an insurance company powered by artificial intelligence, and our experience to date teaches us that artificial intelligence holds enormous promise for the business of insurance. If development of the Principles has focused attention on a values-driven framework for the continual evolution of artificial intelligence, then they already have accomplished much of value.

The clarion call they could be, however, has been obscured by the effects of mutual distrust among the industry, consumers and regulators. Means and ends have become conflated as each faction seeks protection from the others, and the consequent linguistic compromises required to move this project forward have left us with a legalistic document that falls short of giving clear voice to shared values.

In that spirit, we have the following comments on the five Principles:

Fair and Ethical

We believe that artificial intelligence holds within it the power to make the insurance market increasingly more fair and ethical. There would be no reason to object to the principle that AI Actors should know and abide by all applicable legal requirements, including but not limited to Section 4 of the NAIC Unfair Trade Practices Act (Model 880). Similarly, there would be no reason to object to the principle that artificial intelligence

should not further disadvantage already-disadvantaged groups, including groups for which society at large has determined that some measure of redress and/or protection is appropriate.

However, importing specific legal theories into a document avowedly intended to guide development and regulation of a rapidly evolving technology risks stunting that evolution, preventing it from achieving all the good of which it is capable and regulating it inappropriately.

The notion of disparate impact, and its close relative proxy discrimination, come from federal Fair Housing Act jurisprudence. Before prohibiting very specific acts, however, the Fair Housing Act declares a simple and direct principle, "It is the policy of the United States to provide, within constitutional limitations, for fair housing throughout the United States." 42 USCA § 3601. Enforcement also is quite simply stated. In the example of private rights of action:

> An aggrieved person may commence a civil action in an appropriate United States district court or State court not later than 2 years after the occurrence or the termination of an alleged discriminatory housing practice, or the breach of a conciliation agreement entered into under this subchapter, whichever occurs last, to obtain appropriate relief with respect to such discriminatory housing practice or breach.

42 USCA § 3613(a)(1)(A). A discriminatory housing practice is defined in the statute as one of the very specific acts prohibited by the statute. 42 USCA § 3602(f). The prohibited acts do not include proxy discrimination or disparate impact. 42 USCA §§ 3604, 3605, 3606, 3617.

Instead, under regulations of the U.S. Department of Housing and Urban Development as well as case law, the concept of disparate impact is a means of proving a claim of prohibited discriminatory effect. *See* 24 CFR § 100.500; *Texas Dept. of Housing and Community Affairs v. Inclusive Communities Project*, 135 S.Ct. 2507 (2015).

Unlike the *per se* rule in the current version of the Principles ("avoid proxy discrimination against protected classes"), the HUD regulation provides that, if the plaintiff (or the government) proves that a challenged practice caused or predictably will cause a discriminatory effect, then the burden shifts to the defendant who has the opportunity to prove that the challenged practice is necessary to achieve one or more substantial, legitimate, nondiscriminatory interests of the defendant. If the defendant succeeds in proving such necessity, then the plaintiff can still prevail by proving that the interests supporting the challenged practice could be served by another practice that has a less discriminatory effect. 24 CFR § 100.500(c). In declining to provide a safe harbor for the insurance industry, HUD wrote:

> HUD recognizes that risk-based decision making is an important aspect of sound insurance practice, and nothing in the Rule prohibits insurers from making decisions that are in fact risk-based. Under the standard established

2

by the Rule, practices that an insurer can prove are risk-based, and for which no less discriminatory alternative exists, will not give rise to discriminatory effects liability. All the Rule requires is that if an insurer's practices are having a discriminatory effect on its insureds and "an adjustment . . . can still be made that will allow both [parties'] interests to be satisfied," the insurer must make that change.

81 FR 69012-02 (footnotes omitted, quotation in original).

In insurance as well, fairness is a guiding principle aimed at treating similar risks similarly.[1] In artificial intelligence, definition, application and measurement of fairness often are complex. Lemonade believes that AI will evolve toward ever greater granularity in risk categorization and thus toward ever increasing fairness. *See* Schreiber, "AI Can Vanquish Bias" (December 9, 2019)(https://www.lemonade.com/blog/ai-can-vanquish-bias/). To understand why, consider the process of using data to segment a population – that is, to categorize individuals – as evolving in three phases.

In the earliest phase, all people are treated as though they were identical. Everyone represents an average risk and therefore is charged the identical average premium. In this phase, there is no discriminating based on protected classes, but that does not make it fair. In fact, the result is unfair, because the uniform premium is not an actuarially sound estimate of the expected value of all future costs associated with an individual risk transfer.

The next phase sees the population divided into subgroups, according to their risk characteristics. This process is data-driven and facially impartial, but because the data are relatively basic, the groupings are relatively few and crude. This reflects the state of the insurance industry today, and presents two fairness problems.

The first problem is that the groups may be proxies for protected classes. Take the example of gender. Even if women are on average better risks than men, there will be many women who represent better-than-average risks for a woman, and many men who are worse risks than the male average. Treating all men the same and all women the same "unfairly discriminates" against better-than-average women and against average-or-better men. The second problem is that objectively low risk members of group must pay more (per hypothetical "unit" of risk) than their more risky compatriots.

In the third and final phase, crudely-defined subgroups of the population are divided into increasingly homogeneous subgroups. Machine learning does this on a massive scale, using enormous amounts of data, to produce very complex multivariate risk scores. Eventually, each individual will be a "group of one." A grouping that in Phase 2 might have been a proxy

---

[1] The customary standard for insurance rates is that they not be excessive, inadequate or unfairly discriminatory. *E.g.*, NY Ins. Law § 2303. Principle 4 of the Casualty Actuarial Society's Statement of Principles Regarding Property and Casualty Insurance Ratemaking states, "A rate is reasonable and not excessive, inadequate, or unfairly discriminatory if it is an actuarially sound estimate of the expected value of all future costs associated with an individual risk transfer." https://www.casact.org/professionalism/standards/princip/sppcrate.pdf

for a protected class with a risk score of 60, is now viewed instead as individuals, some with a risk score of 90, others of 30 and so forth. This still averages to a risk score of 60, but while that average was imputed to all members of the class in Phase 2, it is applied to none of them in Phase 3. Each person contributes to the pool of funds in direct proportion to the risk they individually represent, rather than their individual share of the risk represented by a large group of somewhat similar people. In this way, artificial intelligence can help avoid the unfairness of rating people in ways that proxy protected classes.

But how can we audit the algorithm to ensure it actually lives up to its promise; how are we to assess the impartiality of its outputs? The answer must lie in their outcomes.

All can readily agree that, if being a woman or a black person resulted in higher premiums regardless of objective risk characteristics, that would be despicable. But a Phase 3 algorithm that identifies an individual's actual riskiness and charges more to assume that risk is fair. The fact that a risk characteristic is unevenly distributed in the population, and more highly concentrated among, say, women means that on average women will pay more. It does not mean that women are charged more *merely because they are women*. As artificial intelligence enables insurance to evolve into Phase 3, the evidence of discriminatory effect will become a different and better measure of disparate impact: differential loss ratios.

If an insurer charges all customers a rate proportionate to the risk they pose, its pure loss ratio should be constant across its whole customer base, but when its customers are reformed into sizable groupings – say by protected class –we should continue to see a uniform pure loss ratio across such cohorts. That would be a signal that even if certain groups are paying more on average, these higher rates are fair by statute and actuarial principle, because they estimate the expected value of all future costs associated with an individual risk transfer. If certain groups exhibit a lower pure loss ratio, that would be a signal that they are being treated unfairly. Their rates are too high, relative to the benefits they are receiving.[2]

Recognizing this helps with another concern about artificial intelligence: that machines can inherit biases. Imagine, for example, that the machine finds that people who are arrested are also more likely to be robbed. An arrest record would, in this hypothetical, become a legitimate factor in assessing property insurance premiums. One problem is that some of the arresting officers may be racist, leading to innocent black people being arrested, the rating algorithm would inherit the humans' racial bias, and a person would pay more for being arrested and consequently more simply for being black.

Machines can overcome the biases that contaminate their training data, if they can continuously calibrate their algorithms against unbiased data like pure loss ratio. Artificial intelligence would quickly determine that having been arrested isn't equally predictive of claims across the population. As data accumulate, the "has been arrested group" would subdivide, because the artificial intelligence would detect that for certain people being

---

[2]   A discriminatory effect could also result from biased claim handling, but a group exhibiting a lower pure loss ratio would signal that as well.

arrested is less predictive of future claims than it is for others. The algorithm would self-correct, adjusting the weighting of this data point to compensate for human bias.

Accountable

Paragraph a of this Principle is heavily redundant with other parts of the Principles,[3] and the concepts embodied in the first two sentences of paragraph a seem more appropriately placed as part of the third Principle. Resolving these drafting issues would result in a Principle better focused on accountability as acknowledgment and assumption of responsibility for actions, decisions and policies including administration, governance and implementation with the obligation to report, explain and be answerable for resulting consequences.

Compliant

As respects privacy laws, we believe this Principle fails to distinguish among types of privacy. There are three types of privacy rights.

Civic Privacy is the right of free citizens to lead their lives as they see fit, safe from interference by their government. Civic Privacy is protected in the "penumbra" of the First, Third, Fourth and Fifth Amendments to the U.S. Constitution. Furthermore, Civic Privacy is held among the liberties that, under the Fifth and Fourteenth Amendments, neither the federal nor any state government may abridge without due process of law.

The Constitutional right of privacy operates to limit solely the actions of government that violate it. Consequently, the concept has limited, if any, application to private commercial artificial intelligence, because the reasons for protecting citizens' privacy from government intrusion are not present in commercial transactions. Private individuals and entities, however, should not become "state actors", aiding and abetting government's abridgement of Civic Privacy. *C.f.*, ECLI:EU:C:2020:559 (U.S.-E.U. Privacy Shield struck down because (i) U.S. law gives U.S. authorities the right to collect personal data about non-U.S. persons without sufficient safeguards and (ii) such individuals have no effective means of redress against the U.S. government in U.S. courts)(Court of Justice of the European Union, July 16, 2020)(http://curia.europa.eu/juris/celex.jsf?celex=62018CJ0311&lang1=en&type=TXT&ancre=).

Social Privacy is the societal norm that people should be able to protect their secrets, control their own identities and regulate the manner of their private interactions with others. It derives from the law of property rights and from community values of "right" and "wrong" social behavior. The concept has little, if any, application to artificial intelligence, except that human beings, in the form of their personal information, should

---

[3]   The second sentence of paragraph a of this Principle is redundant with the first sentence of paragraph a of the first Principle (Fair and Ethical), with the second, third, fourth and fifth sentences of paragraph a of the third Principle (Compliant), the first sentence of paragraph a of the fifth Principle (Secure, Safe and Robust) and to a large extent the first sentence of paragraph b of the fourth Principle (Transparent). The fourth sentence of paragraph a is redundant with the third sentence of the third Principle (Compliant).

not become chattel. In particular, personal information should not be sold without the individual's knowledge and consent.

Commercial Privacy is the ability of one party to a commercial transaction to limit the personal information available to the other party. In an artificial intelligence principle, the notion of commercial privacy ought not to permit insureds to prevent insurers from acquiring and using information concerning the risk they are being asked to assume, provided that insureds have the ability to correct (not fraudulently alter) information obtained from third parties. Furthermore, insurer data collection and use should be logically connected to acceptance and pricing of risk or to investigation and settlement of claims.

Transparent

The final sentence of Paragraph a is altogether too prescriptive, and seems substantively redundant with paragraph b.

Secure, Safe and Robust

The second and third sentences of paragraph a of this Principle seem more appropriately included in the second Principle-Accountable. While it is not altogether clear what effect privacy and unfair discrimination issues could have on security, safety and robustness, it would be appropriate for this Principle to highlight adherence to cybersecurity requirements such as those contained in NAIC Models 668, 670, 672 or 673.

I should like to close by returning to my opening point. Although regulators and regulatory attitudes will have a profound effect (for good or ill) on the evolution of artificial intelligence in the business of insurance, government is conspicuously missing from the definition of "AI Actors." While the document purports to "assist regulators and NAIC committees addressing insurance-specific AI applications," the Principles provide no guidance to regulators, but instead seem to be inflexible standards, grounded in current circumstances, to which insurers will be expected to conform. They do not seem to form a statement of shared principles.

We recommend that mutual distrust be addressed by including regulators in the definition of AI Actors and including a Principal along the lines of the following statement by the Organisation for Economic Co-Operation and Development:

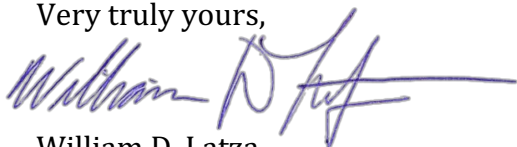> 2.3. Shaping an enabling policy environment for AI
>
>   a)   Governments should promote a policy environment that supports an agile transition from the research and development stage to the deployment and operation stage for trustworthy AI systems. To this effect, they should consider using experimentation to provide a controlled environment in which AI systems can be tested, and scaled-up, as appropriate.

b)   Governments should review and adapt, as appropriate, their policy and regulatory frameworks and assessment mechanisms as they apply to AI systems to encourage innovation and competition for trustworthy AI.

OECD, "Recommendation of the Council on Artificial Intelligence," OECD/LEGAL/0449 (May 22, 2019)(https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449).

Again, thank you for this opportunity to comment. Please do not hesitate to contact me with any questions or if I can provide additional information.

Very truly yours,

William D. Latza
+1 (917) 608-9499
bill.latza@lemonade.com

cc: Daniel Schreiber
        Dennis Monaghan
        Naty Kohn
        Ben Klinger

University of Minnesota Law School

Dear Denise:

I would like to submit a brief comment in connection with the INNOVATION AND TECHNOLOGY (EX) TASK FORCE's consideration of the proposed principles regarding Artificial Intelligence (AI). I understand my comment is late, but I only recently learned of this meeting.

My comment, in short, is that the term "proxy discrimination" has a clear and unambiguous meaning in the insurance setting, notwithstanding comments to the contrary by the insurance industry, In short, proxy discrimination occurs when insurers discriminate based on facially-neutral traits that (i) are correlated with membership in a protected groups, AND (ii) are predictive of losses for precisely that reason. I have written an entire article explaining this point and elaborating on it, which I attach to this email for reference. I would be happy to work with the NAIC and state regulators on this issue, and have been disappointed that no one from the NAIC has reached out to me on this issue despite several attempts on my part to engage on this issue.

https://ilr.law.uiowa.edu/assets/Uploads/ILR-105-3-Prince-Schwarcz-6.pdf

Daniel Schwarcz
Fredrikson & Byron Professor of Law
University of Minnesota Law School
Schwarcz@umn.edu
612-625-4272
Faculty bio: https://www.law.umn.edu/profiles/daniel-schwarcz
Research: http://ssrn.com/author=499486
Twitter: @Dschwarcz