The Market Information Systems Research and Development (D) Working Group recognizes that the term "AI" itself has a variety of contested meanings. Many private sector entities have adopted the term as a marketing concept and apply the label to products simply as a selling point. As such, the term has come to acquire and is an "essentially contested concept."

1

At its most general level, the term "AI" implies machine capacities that mimic or are analogous to processes of human reasoning and learning and entail some degree of machine autonomy in which learning occurs without significant human intervention. Beyond this general description, the Working Group did not feel that an attempt to define the term more strictly would be fruitful. Rather, the term is employed simply as a shorthand reference for a collection of various techniques that algorithmically “general AI,” possess generalized autonomous problem-solving capacities that are comparable to the processes of the human brain, and they are able to adapt to novel situations or information (Macnish et al., 2019).

It is important to emphasize the ways in which AI modeling techniques contrast to the standard scientific model employed in classical or traditional statistics:

<u>Classical Statistics</u>: Method of hypothetical-deductive reasoning in which hypotheses are clearly and narrowly specified prior to data testing, often with a prior understanding of the underlying causal nature of the relationships between variables. Purpose: To further causal understanding.

<u>AI:</u> Often employs a type of “data mining” in which a machine pattern-seeking algorithm is released “into the wild” to identify possible correlations between variables that may be predictive of some independent variable. Hypotheses are not specified prior to data analysis, and the algorithm may very well identify correlations that would not have occurred to an analyst and whose causal relationship is constructed post-hoc (to the degree that AI users are concerned with causality). Purpose: Predict future outcomes or events.

The difference between these two approaches is not trivial, and significant disagreements about the advantages and disadvantages of AI remain. It is of note that AI did not emerge principally from university statistics departments, but rather from the field of computer science. Many statisticians remain skeptical of the techniques and have offered up a variety of caveats for their use. For example, recently the American Statistical Society (ASA) reacted to the “reproducibility crisis” afflicting some disciplines that have discovered, with much consternation, that a large volume of published works could not be replicated. The concern was that increasingly less rigorous statistical methods departing from the hypothetical-deductive approach were becoming more prominent in a variety of fields, undermining confidence on research findings. Remarking on departures from a rigorous hypotheA(f)13 (i)19 regulators and insurers regarding the meaning of statistical relationships appearing in predictive models that lack intuitive or, in many cases, even plausible explanations. See Appendix A for further discussion of the ASA statement.

The discussion above is not intended to sway state insurance regulators one way or the other with respect to AI. The purpose is simply to proffer some caveats shared by many statisticians. A final caveat is the AI techniques were developed to analyze very large data sets consisting of millions of records and possibly thousands or tens of thousands of variables. It is said to have an advantage in that algorithms can perform a large volume of analyses across different constellations of variables in a way that would be highly impractical employing traditional (and manual) model building.

Miscellaneous Data Sources: Some financial data has been incorporated into market information systems. Insurers that are under financial stress, or that rapidly expand into or contract out of a line of business, or that exhibit high defense or other adjudication costs, may be subjected to additional analysis. While financial indicators are only indirect or proxy measures of potential market issues, and by themselves may have no clear market-based interpretation, interpretation within the context of a host of other indicators may be reflective of the present of a market-relevant issue.

The NAIC, in conjunction with state insurance regulators, has developed a broad scope "market score" that incorporates much of the data referenced above, which is made available to regulators via the Market Analysis Prioritization Tool (MAPT). One such data are "normalized" by the premium volume and scope of company operations as necessary. For example, several RIRS-based ratios express the volume of RIRS actions in relation to premium volume, the number of states in which they have significant premium, and a composite ratio that incorporates both premium and scope. Each ratio is given a score, and their contribution to the overall score weighted according to their perceived predictive relevance. For example, financial ratios are accorded significantly less weight than complaints, as their relationship to market misconduct is considered more speculative and indirect.

An important caveat is that predictive analytics is not well developed in market regulation. The ratios employed in the Market Analysis Review System (MARS) have not been subjected to rigorous statistical tests that demonstrate their analytic utility. While some work has been performed in this entities that may merit additional scrutiny and to narrow focus on a much more limited subset of companies out of a larger pool of companies. It therefore primarily prioritizes limited regulatory resources.

State insurance regulators avail themselves of the formal analytical processes adopted by the NAIC. Quantitative or “baseline” analysis identifies entities with anomalous indicators that significantly depart for industry-wide values. A “level 1” analysis may be pursued, in which an analyst devotes additional scrutiny to such things as complaint trends, common reasons complaints are lodged against an insurer, similarities in RIRS actions, etc. If concern still remains (or additional concerns are identified) subsequent to level 1 analysis, a structured level 2 analysis may be performed. A level 2 analysis requires a much greater commitment of time and resources. For example, rather than just manually reviewing complaint data to identify patterns, an analyst may manually review actual complaint documentation to garner a more detailed understanding of the nature of complaints.

As a preliminary to the following discussion, AI/statistical analysis may have two primary functions within the context of the current market analysis structure:

1. More accurately identify companies that merit the additional expenditure of resources necessary to perform the more labor-intensive level 1 and level 2 analyses. Analysis processes that more efficiently identify problem companies for this purpose are by definition more effective and more effectively target resources by avoiding “false positives” (for lack of a better word).
2. Potentially, AI methods could assume many of the functions that are currently performed manually. For example, many of the pattern-seeking analysis performed by analysts in a level 1 review could conceivably be more efficient if automated. Potentially, AI could identify patterns that might elude a human analyst. A very advanced level of AI could perhaps assume complex analysis involved with manually reviewing complaint files and documents. However, while the possibility is raised here, it is not further pursued. That level of AI suitable for tasks may not even exist as yet, or if it does, it may be so specialized that it may not be available to state insurance regulators. Even if available, the likely enormous costs themselves would render them highly impractical.

Whether such AI exists, is available at a practical cost, and can actually outperform more conventional

an14 Td [(it)]TJ (o)-3 (rn6-20 (m)21l)-20 (a)-3 (b)-19 (le)-12 (t)-17 (o)- (b)-19]TJ -07 -0.01da[([(an14 T

approach that forms the core of conventional statistics may have advantages in terms of generating valid causal conclusions. However, AI may have certain advantages with respect to confronting the enormity of modern data. As AI is well-suited to performing much more expansive analysis and pattern-seeking routines over vast quantities of data, it may well identify predictive patterns that would have escaped conventional analysis or that are counterintuitive such that some hypotheses may never have occurred to an analyst employing a standard hypothetical-deductive approach. However, there are distinct disadvantages as well, and they are shared by other approaches often termed "data mining." The fact is that patterns may lack an intuitive meaning, and the manner in which such patterns are identified and render interpretation may be unclear. Additionally, patterns may generate numerous "false positives," apparent patterns or correlations that are purely random and possess no meaning or any real predictive power whatsoever. This is not fatal for AI techniques, but it introduces much in the way of caveats and requires significant remedial measures to be employed. This problem is so significant that it merits a much fuller discussion in a separate section below.

The Work of Market Information Systems Research and Development (D) Working Group

The Working Group solicited input from various parties. Two parties delivered presentations to the Working Group:

1. On June 16, 2021, the Working Group discussed a presentation regarding AI methods currently being explored by NAIC staff to predict which insurers are likely to experience financial stress, including insolvency. Beginning in January 2021, an outside consulting group was retained to develop both AI as well as more traditional statistical techniques to construct predictive models of insolvency risk. The efforts are ongoing at the time of writing. Presenters believed the methods were promising and could mig ( )]TJhect

As noted above, AI techniques such as text analysis could potentially expand such exercises and improve the identification of concerning patterns at a deeper level, as well as assess ways to improve the efficiency of other qualitative tasks.

Recommendation 4: Assess ways AI can improve both the efficiency of qualitative analysis and facilitate pattern recognition across larger volumes of textual evidence, including most especially complaints, but perhaps other textual sources. For example, the “level 1” analysis formalizem-12.1 n7236 0 T are identified via AI and found useful, standard statistical models should also be employed to test whether different techniques yield superior predictive power. Additional discussion of caveats is presented in the appendix.

That said, there is much potential of AI in market analysis, assuming that additional, more granular data are available. As noted, such techniques are most suited for large datasets whose size would make a standard statistical approach impractical just given the sheer number of possible correlations available for testing.

Recommendation 5: Systematically explore potential data sources suitable for AI techniques, with an eye for discovering patterns and relationships in relation to some well-defined outcome one is attempting to predict. This may be identifying entities that may merit additional regulatory scrutiny in a way that is currently done by the less sophisticated methods employed in the MAPT or with the MCAS. Larger volumes of data, such as the standard data requests, can be subjected to AI to identify problematic claims handling, underwriting and other insurer practices.

Summary of Recommendations

Recommendation 1: Survey currently available market analysis data, and identify substantive deficiencies based on the nature and substance of the data elements collected. Ensure that all data are consistently reported across insurers to the degree practical, and ensure adherence to definitions of data elements.

Recommendation 2: In conjunction with recommendation 1 (assess data quality), state insurance regulators should adopt a much more rigorously statistical approach to identify the predictive power of market scoring systems, assess how each variable should be weighted in terms of its unique contribution to productiveness, and drop those that lack analytic utility. In addition, effort should be made to integrate data into a single overall analysis. For example, the MAPT does not incorporate MCAS data, which is typically subject to a separate analysis. The Working Group believes that a "piecemeal" approach is likely less effective than a more integrated approach.

Recommendation 3: In undertaking recommendation 2, incorporate various promising AI modes of analyses, as well as traditional statistical modeling. Constantly assess the precision of model outcomes relative to objectives, such as identifying potential market issues.

Recommendation 4: Assess ways AI can improve both the efficiency of qualitative analysis and facilitate pattern recognition across larger volumes of textual evidence, including most especially complaints, but perhaps other textual sources. For example, the "level 1" analysis formalized in NAIC market system may include a review of the "management discussion and analysis" of the financial annual statement.

Recommendation 5: Systematically explore potential data sources suitable for AI techniques, with an eye for discovering patterns and relationships in relation to some well-defined outcome one is attempting to predict. This may be identifying entities that may merit additional regulatory scrutiny in a way that is currently done by the less sophisticated methods employed in the MAPT or with the MCAS.

AMERICAN STATISTICAL ASSOCIATION RELEASES STATEMENT ON STATISTICAL SIGNIFICANCE AND P-VALUES

Provides Principles to Improve the Conduct and Interpretation of Quantitative Science

March 7, 2016

The American Statistical Association (ASA) has released a “Statement on Statistical Significance and P-Values” with six principles underlying the proper use and interpretation of the p-value [http://amstat.tandfonline.com/doi/abs/10.1080/00031305.2016.1154108#.Vt2XIOaE2MN]. The ASA releases this guidance on p-values to improve the conduct and interpretation of quantitative science and inform the growing emphasis on reproducibility of science research. The statement also notes that the increased quantification of scientific research and a proliferation of large, complex data sets has expanded the scope for statistics and the importance of appropriately chosen techniques, properly conducted analyses, and correct interpretation.

Good statistical practice is an essential component of good scientific practice, the statement observes, and such practice “emphasizes principles of good study design and conduct, a variety of numerical and graphical summaries of data, understanding of the phenomenon under study, interpretation of results in context, complete reporting and proper logical and quantitative understanding of what data summaries mean.”

“The p-value was never intended to be a substitute for scientific reasoning,” said Ron Wasserstein, the ASA’s executive director. “Well-reasoned statistical arguments contain much more than the value of a single number and whether that number exceeds an arbitrary threshold. The ASA statement is intended to steer research into a ‘post $p<0.05$ era.’”

“Over time it appears the p-value has become a gatekeeper for whether work is publishable, at least in some fields,” said Jessica Utts, ASA president. “This apparent editorial bias leads to the ‘file-drawer effect,’ in which research with statistically significant outcomes are much more likely to get published, while other work that might well be just as important scientifically is never seen in print. It also leads to practices called by such names as ‘p-hacking’ and ‘data dredging’ that emphasize the search for small p-values over other statistical and scientific reasoning.”

The statement’s six principles, many of which address misconceptions and misuse of the p-value, are the following:

1. P-values can indicate how incompatible the data are with a specified statistical model.
2. P-values do not measure the probability that the studied hypothesis is true, or the probability that the data were produced by random chance alone.
3. Scientific conclusions and business or policy decisions should not be based only on whether a p-value passes a specific threshold.