

---

# Big data in insurance contracts – a tool for good, or bad?

---

**Michele van Eck**

*BCom (Law), LLB, LLM, LLD, BTh, BTh(Hon)*

*Associate Professor, Department of Private Law, University of Johannesburg*

**Samantha Huneberg**

*BCom (Law), LLB, LLM, LLD*

*Senior Lecturer, Department of Mercantile Law, University of Johannesburg*

## SUMMARY

Big data is changing the way many companies conduct business on a day-to-day basis. Insurers are notorious for utilising data in the risk assessment of prospective and current policyholders. The use of such risk assessment mechanisms in insurance has resulted in some discrimination between various policyholders but this has been held to be justifiable due to the fact that it is based on actuarial science and is therefore viewed as fair. However, the advent of big data, data analytics, algorithms, and artificial intelligence is providing insurers with far more sophisticated data about potential policyholders. This may prove to be beneficial to insurers in many ways, but it also brings about additional possibilities of exclusion within the industry. This use of big data has the potential to increase social injustices within our country, which is something that needs to be avoided as much as possible. Discrimination, through the use of big data, is a reality that needs to be addressed by insurers and regulators alike. Achieving social justice as far as possible in the insurance industry is crucial and also requires considerations in the areas of morality and ethics. The reason for this is that the very nature of big data is integrally linked to the assessment of the policyholder's moral risks and hazards for the benefit of the insurer is often linked to personal circumstances and, sometimes, the financial circumstances of the policyholder, and may even speak to the policyholder's integrity. Although potentially beneficial for the insurers from a risk assessment perspective, the use of big data has ethical and moral considerations within the insurance context. After all, the insurance industry's collection, storage, and use of big data raises ethical and moral concerns and casts a shadow on the manner it is used to assess policyholders. This discussion highlights the need for regulatory oversight of big data, an aspect that is notably missing in the South African legislative framework.

## 1 Introduction

The underlying relationship between the insurer and insured is contractual and therefore, the general requirements of the law of contract form the basis of an insurance policy.<sup>1</sup> Insurers have always made use of data in the assessment of the policyholder's risk (also called risk classification),<sup>2</sup> which may include moral risks and hazards of the policyholder.<sup>3</sup> It is trite that data and information have been, and continue to be, the basis of the assessment of risks and consequently the provision of insurance.<sup>4</sup> Bhoola, Peick, Pio and Tshabalala reiterate this notion and state that "the industry is built on the capabilities of analysing data to understand and evaluate risks".<sup>5</sup> Further, at the inception of the modern insurance industry, the development of the actuarial and underwriting professions largely depended upon the use of data analytics.<sup>6</sup> Therefore, the ability to either obtain or store data of the policyholder for the continued analysis of risk prior to and for the duration of the insurance contract is of great importance to the insurer. Although generally accepted that such risk classifications would result in some discrimination within the insurance industry, it has been long held that the differentiation between policyholders would be justifiable due to it being backed by actuarial data.

Risk classification has become more sophisticated as insurers are today making increased use of big data, data analytics, algorithms, and artificial intelligence (AI) to provide more personalised cover and appropriate insurance products to policyholders.<sup>7</sup> These technological innovations, specifically the use of big data, are beneficial for the insurance industry but there remains an element of risk of financial

---

1 Reinecke, van Niekerk and Nienaber *South African Insurance Law* (2013) 111.

2 For the purpose of this article the term "policyholder" refers to both the potential policyholder as well as an existing policyholder in an insurance contract.

3 See, for example, *Munns and Another v Santam Ltd* 2000 4 SA 359 (D) para 64.

4 Bhoola, Peick, Pio and Tshabalala "Big Data Analytics" 2014 114 <https://actuariesociety.org.za/convention/convention2014/assets/pdf/papers/2014%20ASSA%20Bhoola%20Kruger.pdf> (last accessed 2022-03-02). For decades, insurers have relied heavily on data in order to assess consumers risk selection (pricing and underwriting), marketing (distribution and sales activities), product development and even claims management (fraud detection). This traditional use of data generally included consumer demographics, medical information, as well as behavioural and loss data. This data was generally obtained from the policyholder themselves. See also, for example, *Potocnik v Mutual and Federal Insurance Co Ltd* 2003 6 SA 559 (SE).

5 Bhoola, Peick, Pio and Tshabalala 2014 114.

6 Bhoola, Peick, Pio and Tshabalala 2014 114-115.

7 BEUC "The use of big data and artificial intelligence in insurance" 2020 2. [beuc-x-2020-039\\_beuc\\_position\\_paper\\_big\\_data\\_and\\_ai\\_in\\_insurances.pdf](https://www.beuc.eu/media/14472/2020-039-beuc_position_paper_big_data_and_ai_in_insurances.pdf) (last accessed 2022-03-02).

exclusion for policyholders<sup>8</sup> and consequently perpetuating social injustices. This is especially relevant for countries like South Africa where financial exclusion remains a fundamental issue for most of the population.<sup>9</sup> Other risks in the use of big data in the insurance industry include the protection of policyholder's information and privacy as well as ensuring that the information of the policyholder is used reasonably in assessing the insurer's risks in insurance contracts.<sup>10</sup>

There are certainly justifiable apprehensions about the insurance industry's use of big data, data analytics, algorithms, and AI due to the possibility of direct or indirect discrimination that it presents to policyholders.<sup>11</sup> Further to this, achieving social justice within the insurance industry also requires the consideration of subjective elements of morality and ethics. After all, the nature of big data is integrally linked to the assessment of the moral risks and hazards of policyholders for the benefit of the insurer that is often linked to personal circumstances and, on occasion, the financial circumstances of the policyholder,<sup>12</sup> and may even speak to the policyholder's integrity.<sup>13</sup> Although big data may present information to the insurer to assess the moral risks and hazards of the policyholder, the use of big data has, in itself, ethical and moral considerations within the context of insurance. After all, the insurance industry's collection, storage and use of big data raises ethical and moral concerns and casts a shadow on the manner it is used to assess policyholders.

It is against this background that this article intends to consider how big data, data analytics, algorithms, and AI have highlighted pre-existing practices that may, if left unregulated, expand the fissures of social inequalities already prevalent in South African society.

## 2 Data, information, knowledge, and algorithms

The term "data" derives from its singular Latin form "datum",<sup>14</sup> however, generally it has become common practice that the term data

---

8 Note the comments in *Munns and Another v Santam Ltd* 367, wherein certain circumstances a person's financial position may link to the moral risk and hazard of the policyholder.

9 Chitimira and Ncube "The Role of Regulatory bodies and other Role players in the Promotion of Financial Inclusion in South Africa" 2020 *AUDJ* 7.

10 The issue of privacy in the use of big data is a significant factor in its use. Some protections have been afforded in the promulgation of the Protection of Personal Information Act 4 of 2013 (POPIA). The discussion of POPIA falls outside the scope of this article.

11 Richards & King "Big Data Ethics" 2014 *Wake Forest Law Review* 408.

12 See also *Munns and Another v Santam Ltd* 367.

13 *Commercial Union Insurance Co of SA Ltd v Wallace NO; Santam Insurance Ltd v Africa Addressing (Pty) Ltd* 2003 JOL 11584 (SCA) 31-32.

14 Merriam-Webster Dictionary entry "datum" <https://www.merriam-webster.com/dictionary/datum> (last accessed 2022-04-16).

may be used for both singular and plural forms.<sup>15</sup> Data has become the backbone of society.<sup>16</sup> Clegg mentions that the full potential of data has been limited as a result of technological limitations in analysing the data. The Fourth Industrial Revolution has changed this limitation and has unlocked the “latent power” of data in the insurance industry.<sup>17</sup> This notwithstanding, data is just part of the equation – after all, data without analyses or context is often meaningless.

According to Clegg, the manner in which data is understood can be represented in a pyramid, wherein the raw data is at the base of the pyramid.<sup>18</sup> Such data may be described as “factual information (such as measurements or statistics) used as a basis for reasoning, discussion, or calculation”<sup>19</sup> or “information in digital form that can be transmitted or processed”.<sup>20</sup> Resting upon data, being the base of the pyramid, is the presentation and construction of information which brings meaning to the raw data.<sup>21</sup> Finally, at the apex of the pyramid is what Clegg calls “knowledge”, which is the understanding wherein the information is interpreted and analysed to give it meaning and context to be useful.<sup>22</sup> This may also be referred to as structuring information.

The ability to structure, interpret and contextualise large volumes has always been a challenge, but with the advent of the modern computer, the task was less daunting. The modern age of technology has brought about vast amounts of data and information. In order to structure, interpret, and contextualise information into something useful in the digital age, algorithms are used, which are computer programmes that involve “a step-by-step procedure for solving a problem or accomplishing some end”.<sup>23</sup> Evidently, these algorithms give rise to sophisticated data analytics that form the basis of so-called big data. Technological developments in the Fourth Industrial Revolution utilise data and algorithms to provide processes that enable the collection and structuring of such data.

### 3 What is big data?

The Fourth Industrial Revolution has changed the operation of the insurance industry. Dolman, Frees, and Huang note that “insurers are

---

15 Clegg *Big Data How the Information Revolution is Transforming Our Lives* (2017) 7.

16 Clegg (2017) 4.

17 Clegg (2017) 4-5.

18 Clegg (2017) 7.

19 Merriam-Webster Dictionary entry “data” <https://www.merriam-webster.com/dictionary/data> (last accessed 2022-04-16).

20 Merriam-Webster Dictionary entry “data” <https://www.merriam-webster.com/dictionary/data> (last accessed 2022-04-16).

21 Clegg (2017) 7.

22 As above.

23 Merriam-Webster Dictionary entry “algorithm” <https://www.merriam-webster.com/dictionary/algorithm#note-1> (last accessed 2022-04-16).

redefining the way that they do business with the increasing capacity and computational abilities of computers, availability of new and innovative sources of data, and advanced artificial intelligence algorithms that can detect patterns in data that were previously unknown”.<sup>24</sup> The insurer’s increased utilisation of algorithms and advanced data analytics,<sup>25</sup> as well as the larger availability of traditional<sup>26</sup> and non-traditional data<sup>27</sup> sources, is set to change the way the traditional insurance industry operates.<sup>28</sup> One of these fundamental changes is the availability and use of data.

The term “big data” does not have an unequivocal definition and a universally accepted definition is somewhat elusive.<sup>29</sup> Big data may be described as the process of collating significant amounts of information from various sources in order to provide “decision-making insights”.<sup>30</sup> Therein, big data comprises a process that results in information ultimately encouraging decision-making.<sup>31</sup> It has been suggested that big data “may extract new insights or create new forms of value, in ways that change markets ...”.<sup>32</sup> As big data draws information from all sectors, including areas of the policyholder’s life and social media platforms, it can create a relatively accurate representation of the policyholder and create certain predictions of the character and behaviour of the policyholder. To do so, algorithms are used to assess and contextualise data.

The algorithms employed within big data provide endless possibilities for the insurers which include direct policyholder servicing, such as

- 
- 24 Dolman, Frees and Huang “Multidisciplinary collaboration on discrimination – not just ‘Nice to Have’” 2021 *Annals of Actuarial Science* 485.
- 25 In the paper this will be collectively referred to as “big data analytics” (BDA).
- 26 Traditional data is also known as structured data which refers to data that resides in a fixed field within a record. It includes data contained in business process applications utilising databases and spreadsheets. Structured data is more organised data that is usually simpler to manage and analyse. It can also be stored in sources such as Excel, Access, and other traditional software. See Bhoola, Peick, Pio and Tshabalala 2014 108.
- 27 Non-traditional data is also known as unstructured data. Examples include unformatted paper-based documents, streaming instrument data, machine logs, webpages, pdf files, PowerPoint presentations, emails, blog entries, wikis and word processing documents. See Bhoola, Peick, Pio and Tshabalala 2014 108.
- 28 International Association of Insurance Supervisors (IAIS) “Issues paper on the use of big data analytics in insurance” 2020 7 <https://www.iaisweb.org/file/89244/issues-paper-on-use-of-big-data-analytics-in-insurance> (last accessed 2022-04-16).
- 29 Richards & King 2014 *Wake Forest Law Review* 397; see also OECD “The impact of big data and artificial intelligence (AI) in the insurance sector” 2020 10 [www.oecd.org/finance/impact-big-data-ai-in-the-insurance-sector.htm](http://www.oecd.org/finance/impact-big-data-ai-in-the-insurance-sector.htm) (last accessed 2022-04-16).
- 30 Kalyvas and Overly *Big Data: A Business and Legal Guide* (2015) 1.
- 31 As above.
- 32 Mayer-Schonberger & Cukier *Big Data: A Revolution that will transform how we live, work and think* (2013) 6.

providing automated advice and pre- and post-sales support, as well as improved claims handling.<sup>33</sup> Ancillary to the above, the algorithms can assist in designing targeted advertising campaigns for prospective policyholders, they can also serve in obtaining specific insights on policyholder preferences, as well as influence policyholder behaviour.<sup>34</sup> Furthermore, big data can inform insurance product design<sup>35</sup>, determine policyholders credit history, assist in risk selection and pricing of products, the algorithms can also assist insurers in getting to know their customers and their needs.<sup>36</sup> These benefits can go a long way in providing efficient and suitable products and services to policyholders. Therein, big data is an excellent tool to assess the moral risks and hazards of the policyholder. In other words, the algorithms that assess big data draw a picture of the policyholder's integrity and moral profile to assess relevant risks in issuing an insurance contract.

Big data is made up of structured,<sup>37</sup> unstructured data,<sup>38</sup> as well as semi-structured data, which integrates the fixed format of structured data and the varied format of unstructured data but importantly does not correspond primarily to either data format. The term is often defined in terms of five characteristics, specifically; volume, velocity, variety, veracity, and value.<sup>39</sup>

- The first characteristic of “volume” refers to the size of the data collated.<sup>40</sup> The increase in the volume of data available for analysis and the growth in data storage is attributable to new forms of data that were not previously the focus of structured data analytics, namely, audio, video, and large media on social media platforms.<sup>41</sup>
- Second, “velocity” refers to the speed of data generation and processing.<sup>42</sup> Velocity considers the pace of the flow of data from various sources such as processors, machines, and other platforms, such as social media.<sup>43</sup> The volume of available data has altered how data is considered and analysed.
- Third, “variety” refers to the different forms and range of data available for processing.<sup>44</sup> Data may comprise numerous forms which were not

---

33 IAIS Issues Paper 2020 11.

34 For example, how Discovery influences their customers to buy healthier foods by getting money back as well as driving better and getting money back; see IAIS Issues Paper 2020 11.

35 For example, like switching insurance products on and off based on use.

36 IAIS Issues paper 2020 11.

37 OECD 2020 10. Structured data refers to data in tables and defined fields; see also n 26 above.

38 OECD 2020 10. Unstructured data refers to things such as social media postings, reports and recorded interviews as well as pictures such as satellite imaging; see n 27 above.

39 Pagallo “The legal challenges of big data: Putting secondary rules first in the field of EU data protection” 2017 *EDPL* 36.

40 As above.

41 Bhoola, Peick, Pio and Tshabalala 2014 108.

42 Pagallo 2017 *EDPL* 36.

43 Bhoola, Peick, Pio and Tshabalala 2014 109.

44 Pagallo 2017 *EDPL* 36.

previously employed in the processing stage.<sup>45</sup> Whilst this variety of data provides increased decisional insight, the variety of available data presents unique challenges for mining, storage, and analysis of data.<sup>46</sup>

- Fourth, “veracity” refers to user entry errors, redundancy, and/or potential for corruption of the data being processed.<sup>47</sup> Veracity pertains to the confidence in and accuracy of the data being processed.<sup>48</sup>
- Fifth, “value” refers to the analytical findings being both insightful and practical in application.<sup>49</sup> Fundamentally valuable data may be utilised by organisations to solve business challenges, such as what we are seeing in the insurance industry.<sup>50</sup>

Big data may impact the insurance industry and the entire product lifecycle, as it changes how insurers do business.<sup>51</sup> Yet, there is a plethora of benefits posed using big data in the insurance industry. If insurers utilise big data, then this may positively impact various phases of the product life cycle, all the way from the underwriting phase,<sup>52</sup> to marketing, and even in the claims management phase.<sup>53</sup> Big data permits the insurer to consider all the available data to make an informed decision regarding the risk of insuring the policyholder.<sup>54</sup> There are inherent risks that the use of big data may result in insurance contracts either being concluded, or insurance being refused without human intervention or consideration. This may be open to abuse and raises the risk of discrimination and ethical considerations, which are discussed later in this article. However, the challenge for insurance regulators is to examine whether the use of big data may ultimately lead to unfair and discriminatory practices for potential and existing policyholders. This would have to be assessed both in the pre-contractual phase<sup>55</sup> where the

---

45 Such as email, video, audio, photographs, and monitoring devices; See Bhoola, Peick, Pio and Tshabalala 2014 109.

46 Bhoola, Peick, Pio and Tshabalala 2014 109.

47 Pagallo 2017 *EDPL* 36.

48 The processor must be able to trust the data for utilisation in decision-making.

49 Bhoola, Peick, Pio and Tshabalala 2014 110.

50 As above.

51 Frees and Huang “The Discriminating (Pricing) Actuary” *North American Actuarial Journal*, (2021) 1 14. The insurance product lifecycle refers to the various phases of an insurance product. The life cycle includes: product design and development, promotion and marketing, advice, point of sale, information after point of sale and claims and complaint handling; see Millard “The impact of the twin peaks model on the insurance industry” 2016 *PELJ* 29.

52 Underwriting in insurance refers to upfront underwriting which means evaluating the client and the risks insurers are willing to insure.

53 OECD 2020 11.

54 See also, the example in *Trust Bank van Afrika Bpk v President Versekeringsmaatskappy Bpk en 'n Ander* 1988 1 SA 546 (W), wherein the overall structure of the organisation was to be considered in order to assess the moral risk of the policyholder.

55 The pre-contractual phase in insurance refers to the period before the policy is concluded between the insurer and insured. It is therefore referring to the negotiation stage between the parties. See Reinecke, van Niekerk and Nienaber (2013) 95.

risks of providing insurance to the potential policyholder are assessed, as well as in the post-contractual phase<sup>56</sup> wherein the risk of the policyholder is assessed for the continued provision of insurance.

## 4 A dark side to big data

As with any new technological advancement, there is always the potential for unknown challenges and risks. The insurer's use of big data and a general increased reliance on these systems for the processing of customer data generates several potential challenges and risks.<sup>57</sup> Big data does not necessarily bring about new discriminatory practices in the insurance industry, but rather, according to Dolman, Frees, and Huang the challenges and discriminatory practices that were in existence in the insurance industry before the use of big data.<sup>58</sup>

The first challenge is with technology itself. Algorithms, due to their very nature, are usually complex, and it is due to this complexity that they are often considered highly confidential.<sup>59</sup> This confidentiality can, therefore, lead to a lack of transparency and severe asymmetry of understanding and information between the people who design and use the algorithms, as well as policyholders and insurers seeking to understand the results produced by these algorithms.<sup>60</sup> If algorithms are not designed to create fair, or at the very least reasonable, results (which is impossible to determine without understanding the nature and content of the proprietary code), it may lead to stereotyping and consequently discriminatory results.<sup>61</sup> Further to that, an algorithm's effectiveness is dependent on the quality, completeness, and accuracy of the available data. Therefore, it is possible that the data may be impeded by potential errors in its preliminary design or programming. Adding to the complications, some algorithms are based on machine learning, meaning "that as the algorithm gathers and analyses more data, it can modify itself without human intervention after initiation".<sup>62</sup> These specific machine learning-type algorithms may raise questions regarding the accountability and transparency of the decision-making process.<sup>63</sup> Unfortunately, some studies illustrate that algorithms inherent in the bias

---

56 The post-contractual phase in insurance refers to the period after that policy has been concluded between the insurer and insured. It generally refers to their dealings after the point of sale and would include claims handling and complaints management.

57 Being the algorithms.

58 Dolman, Frees and Huang 2021 *Annals of Actuarial Science* 485.

59 IAIS Issues Paper 2020 11.

60 As above. Clear documentation should be utilised as it may assist to ensure improved transparency and understanding.

61 Dolman, Frees and Huang 2021 *Annals of Actuarial Science* 485.

62 IAIS Issues Paper 2020 11.

63 OECD 2020 18. Take note that developments in machine learning capabilities may even result in the operation of algorithms becoming more complex, even for those who design them. See also IAIS Issues Paper 2020 11.

of the programmers,<sup>64</sup> may result in discriminatory practices and stereotyping. Therefore, the use of algorithms may negatively impact social justice within society.

The use of big data also brings certain ethical factors into consideration, or what some have called a “moral obligation” as to how big data is used.<sup>65</sup> One such ethical consideration is whether individualisation of insurance policies, relying on big data, and advanced analytics is fair and whether it takes into consideration uncontrollable risks.<sup>66</sup> A controllable risk is a risk that can be changed based on applicable mitigation measures. For example, promoting better driving behaviour or reducing one’s speed. Insurers, like Discovery, are relying on big data of their consumers to incentivise them to change their behaviour when it comes to shopping and exercising, and this may be viewed as intrusive and manipulative by consumers.<sup>67</sup> In comparison, an uncontrollable risk is something that is outside the policyholder’s control. This generally relates to the location of the policyholder’s home in a natural disaster-prone area or, a genetic predisposition as far as life insurance is concerned. These examples can lead to a policyholder being deemed to be a higher risk, which can then ultimately lead to the insurance coverage being completely unaffordable or even worse, being completely excluded from coverage altogether.<sup>68</sup>

The possibility of inequitable and unsuitable treatment of the policyholder is also a potential risk. The reason for this is that machine learning algorithms are based on historical data and therefore, commonly replicate the past, so therefore there may be an increase in the likelihood that these algorithms perpetuate unforeseen biases.<sup>69</sup> This highlights the possibility of discrimination (the act of treating one group of people differently from another),<sup>70</sup> and although the potential for biased decision-making has always existed in the insurance industry and is not unique to algorithms, it is clear that self-learning algorithms can act on a much broader scale and at a faster pace than the previous regime.<sup>71</sup> This highlights the possibility of unfair discrimination that big data brings to the insurance industry, which is discussed further below.

---

64 See, for example, Frees and Huang 2021 *North American Actuarial Journal* 14.

65 Zwitter “Big Data ethics” 2014 *Big Data & Society* July–December 2.

66 Huneberg and Berry “Big data and the paradigmatic shift from passive insurer to risk personalisation” 2022 *TSAR* 300.

67 IAIS Issues Paper 2020 12. Strong consumer rights are a necessary precondition to minimise the potential risks associated with these digital transformations and to ensure that consumers and society as a whole can benefit from these innovations.

68 IAIS Issues Paper 2020 12.

69 IAIS Issues Paper 2020 11-12.

70 Definition adapted from Dolman, Frees and Huang 2021 *Annals of Actuarial Science* 485.

71 IAIS Issues Paper 2020 11-12.

## 5 Big data as a tool for discrimination

### 5.1 Introductory comments

When taking out insurance, consumers should have a right not to be discriminated against,<sup>72</sup> yet at the same time, the insurance industry is based on some form of discrimination, which has often been argued as being justified in the nature of insurance.<sup>73</sup> Frees and Huang describe it as follows:<sup>74</sup>

Insurers collect information on current and potential customers. They collect information about the customers themselves, the entity being insured (whether a person, organization, or physical object such [as] an auto or home), where the entity is located (that can vary, such as a person or auto), and parameters about the contract desired, among other things. This information, represented as variables or factors, provides the basis that insurers use to form groups and make decisions. By treating groups differently, they discriminate among them.

Such discriminatory practices may occur at different touchpoints in the insurance life cycle. First, it may occur in deciding whether or not insurance must be provided to the policyholder or not, which would be at issuing insurance or at the renewal of the insurance policy.<sup>75</sup> Second, the insurer may limit the scope or application of the insurance.<sup>76</sup> Finally, the price of premiums may also be adapted and changed depending on the data at the insurer's disposal.<sup>77</sup> Therefore, it can be said that the insurance industry is no stranger to the differential treatment of its customers, but unfair discrimination is different and must be avoided.<sup>78</sup> Kuschke states that the very nature of insurance is based on discrimination.<sup>79</sup> Further, she states that "persons who pose a higher risk or chance of loss pay higher premiums than others for the same insurance cover, and stereotypes are used to predict insurance risks. Whether this discrimination is fair or not, depends on constitutional norms".<sup>80</sup> What is considered justifiable in the insurance industry is continuously evolving. Take for example, in 2012 the European Court of Justice provided a preliminary ruling in the matter of *Association belge des Consommateurs Test-Achats ASBL, Vann van Vugt, Charles Basselier v Conseil des ministres* Case C-236/09 ECJ where it was ruled that the

---

72 BEUC 2020 11.

73 Frees and Huang 2021 *North American Actuarial Journal* 1.

74 Frees and Huang 2021 *North American Actuarial Journal* 2.

75 As above.

76 As above.

77 As above.

78 See *Minister of Finance v Van Heerden* 2004 6 SA 121 (CC).

79 Kuschke "Association belge des Consommateurs Test-Achats ASBL, Vann van Vugt, Charles Basselier v Conseil des ministres Case C-236/09 ECJ" 2021 *De Jure* 626.

80 As above.

differentiation between men and women in the provision of insurance was an unacceptable practice.<sup>81</sup>

The use of big data and algorithms may perpetuate unlawful or unfair discrimination, or at the very least, highlight discriminatory practices that already exist within the insurance industry. The increased use of algorithms and big data in insurance carries the risk of unfair discrimination and potential financial exclusion of consumers.<sup>82</sup> This is particularly the case when the differentiation between policyholders shifts from an economic decision-making tool to something more nefarious based on a person's moral or social standing in society.<sup>83</sup> Generally, anti-discrimination laws are concerned with formal or intentional discrimination. The existing laws do not provide much guidance on algorithmic decision making and that is because typical discrimination focuses on the human component of the decision.<sup>84</sup>

## 5 2 South African legislative framework

As a starting point, the preamble of the Constitution of the Republic of South Africa 1996 notes that the purpose of the Constitution is to “heal the divisions of the past and establish a society based on democratic values, social justice, and fundamental human rights”. Therefore, all legislative engagements must reinforce this constitutional imperative. Section 9 of the Constitution also refers to the term “unfair discrimination” as opposed to some countries that only refer to the term “discrimination”.<sup>85</sup> Unfair discrimination is described as when a person is treated differently as compared to other categories of people and therefore a person's dignity as a human being is impaired by such treatment. Section 9(4) states that “no person may unfairly discriminate directly or indirectly against anyone on one or more [prohibited] grounds ...”. The prohibited grounds listed in the Constitution include race, gender, sex, pregnancy, ethnic or social origin, colour, sexual orientation, age, disability, religion, conscience, belief, culture, language, and birth, etc.<sup>86</sup> However, the Constitution does not directly apply to contractual engagements, but would, at best, only apply indirectly to insurance contracts by means of the application of public policy considerations.<sup>87</sup>

81 See further discussion on this in Kuschke 2012 *De Jure* 624 – 630. In South Africa, see the case of *Robert v Minister of Social Development* Case Nr 32838/ 05 TPD for an evaluation of gender discrimination in the context of social grants.

82 As above.

83 Adapted from Frees and Huang 2021 *North American Actuarial Journal* 2.

84 Gillis and Speiss “Big Data and Discrimination” 2019 *The University of Chicago Law Review* 462.

85 S9 of Chapter 2 of the Constitution, 1996; see SAIA SAIA supports the traditional insurance practice of differentiation in insurance premiums to the benefit of consumers (fanews.co.za) 2011 (last accessed 2022-04-16).

86 S9(3) of Chapter 2 of the Constitution, 1996.

87 See *Barkhuizen v Napier* 2007 5 SA 323 (CC); *Beadica 231 CC v Trustees, Oregon Trust* 2020 5 SA 247 (CC).

The Promotion of Equality and Prevention of Unfair Discrimination Act (PEPUDA) gives effect to the Constitutional imperative that no person shall be unfairly discriminated against<sup>88</sup>, and prohibits unfair discrimination, harassment and hate speech. Section 14 of the PEPUDA states that in order to defend an action of unfair discrimination, the said discrimination should reasonably and justifiably differentiate between persons according to objectively determinable criteria, intrinsic to the activity concerned and that would amount to a valid defence against a claim of unfair discrimination. However, PEPUDA does not prohibit all forms of discrimination but rather only prohibits unfair discrimination. There are certain circumstances where discrimination can be regarded as fair.<sup>89</sup> In determining whether discrimination is fair, account may be taken of whether the discrimination “reasonably and justifiably differentiates between persons according to objectively determinable criteria (for example, actuarial evidence), intrinsic to the activity concerned (for example, insurance business)”.<sup>90</sup> However, there have been developments herein with the matter of *Association belge des Consommateurs Test-Achats ASBL, Vann van Vugt, Charles Basselier v Conseil des ministres* Case C-236/09 ECJ,<sup>91</sup> that provided the preliminary ruling that differentiating between male and females in insurance would be considered to be gender discrimination. In the South African insurance industry, the differentiation between male and female is still considered to be an acceptable practice.

This notwithstanding, the South African insurance industry has always concerned itself with the fact that if an insurer can show with statistics, over a particular period of time, that the claim ratios confirm an increase in and exposure to risk as a result of age, gender, location or other factors,<sup>92</sup> premiums will be accordingly determined.<sup>93</sup> It is trite that the differentiation in premiums is an actuarial science and does not meet the threshold of unfair discrimination.<sup>94</sup> Differential treatment, in this specific context, is not viewed as unfair discrimination due to the actuarial data used in support of such decision making. However, big data may change this narrative, due to its capabilities which go beyond the current forms of risk determination and differentiation. There is, therefore, an increased risk of unfair discrimination of policyholders in

---

88 4 of 2000.

89 For example, measures designed to advance persons disadvantaged by the previous system of racial discrimination.

90 S6 of the Equality Act. Read with S14(2)(b). S6 of the Equality Act prohibits unfair discrimination in general. To prove that the discrimination is fair, one must take into account whether the discrimination reasonably and justifiably differentiates between persons according to objectively determinable criteria that are intrinsic to the activity concerned. Further, Part 5 of the Schedule to the Equality Act specifically refers to discriminatory practices in insurance services.

91 See further discussion on this in Kuschke 2012 *De Jure* 624 – 630.

92 Including that of moral risks as was the case in *Potocnik v Mutual and Federal Insurance Co Ltd*.

93 SAIA 2011.

94 As above.

the application of big data which requires regulation. There are generally three forms or areas in which discriminatory practices may develop in the use of big data in the insurance industry, being that of the policyholder, the price of the premium, and the data itself. These are discussed further in the paragraphs that follow.

### 5.3 First example of discrimination: Personalisation of offerings

The assessment of personalised insurance for policyholders stems from their individual risk profiles. Increasingly powerful algorithms may tempt insurers to collect a wider array of data about policyholders,<sup>95</sup> and will give insurers insights into policyholder's future conduct and the likelihood of a customer making a claim.<sup>96</sup> The predictions in this context may result in hyper personalised insurance cover, although advantageous in certain ways, may result in some policyholders becoming completely 'uninsurable', based on the risk assessments offered by big data. Increased personalisation and targeting of insurance products can possibly lead to discriminatory results for consumers. A major problem arises when other types of data points could practicably act as proxies for possible discriminatory traits or could closely be correlated with protected characteristics such as race and/or gender<sup>97</sup> and lead to proxy discrimination instead.<sup>98</sup> It may also negatively impact the benefits for the policyholder in that of insurance risk pooling.<sup>99</sup> The fact that AI systems can learn from data does not guarantee that their outputs will be free of human bias or discrimination, and there is plenty of evidence of AI systems picking up existing human biases or historic discrimination.<sup>100</sup> This is clearly a potentially discriminatory practice,<sup>101</sup> which may in certain instances meet the threshold of "unfair discrimination" and have the effect of stereotyping the policyholder and the policyholder's lifestyle. Basing a decision to insure or not to insure a policyholder on this alone, without human intervention or the application of common sense, may feed into pre-existing bias and stereotypes in the application and use of such algorithms.

Unfortunately, most types of discrimination generated by big data cannot be dealt with using current anti-discrimination laws, as these laws traditionally focus on discrimination based on protected characteristics, such as race, by humans.<sup>102</sup> Big data can use classes and categories for differentiation that do not directly relate to protected characteristics.<sup>103</sup>

---

95 BEUC 2020 3.

96 As above.

97 For instance, postcodes signalling ethnicity or occupation categories signalling gender.

98 BEUC 2020 11.

99 Frees and Huang 2021 *North American Actuarial Journal* 14.

100 BEUC 2020 12.

101 BEUC 2020 3.

102 As above.

103 As above.

There will be a greater need for regulatory safeguards if the risk of proxy discrimination rises.<sup>104</sup> Supervisors may need, for example, powers to eliminate the use of certain data points that are unnecessary or could be potential sources of biases, and regularly audit algorithms in order to detect potentially unlawful discriminatory outcomes.<sup>105</sup> Furthermore, insurers should thoroughly test big data algorithms prior to their release to ensure that there are no possible discriminatory results.<sup>106</sup>

#### **5 4 Second example of discrimination: Personalisation of pricing**

Frees and Huang explain that the determination of insurance pricing stems from the 17th-century theory of randomness<sup>107</sup> and that fairness in this context was based on a “framework of individual equitable contracts, ones that traded a certain present amount for an uncertain future value”.<sup>108</sup> This theory of randomness became the backbone to insurance pricing in the 19<sup>th</sup> century,<sup>109</sup> wherein a premium had to be paid for “compensation for an uncertain future event”.<sup>110</sup> This can be described as insurance pooling and a way to address uncertainties and losses for insurers.<sup>111</sup> Risks associated with the insurance pools depend on the nature of the policyholders and may impact the premium and insurance costs accordingly.<sup>112</sup> Discriminatory price optimisation practices can occur through the employment of Big Data as they will allow insurers to charge more individualised prices.<sup>113</sup> Therefore, insurers relying on big data analytics will be able to charge differential prices to groups of consumers, which will then lead to more personalised pricing based on the behavioural characteristics and personal data of consumers.<sup>114</sup>

While this type of personalised pricing may bring benefits to certain policyholders, there is developing evidence of pricing practices in the insurance industry that fail to treat policyholders fairly, because insurers are partaking in harmful price optimisation techniques.<sup>115</sup>

Evidence suggests that insurers may “charge prices based on the optimum amount of margin they can earn from an individual policyholder, rather than the risk and/or cost of the individual

---

104 As above.

105 BEUC 2020 13.

106 As above.

107 Frees and Huang 2021 *North American Actuarial Journal* 3.

108 As above.

109 As above.

110 As above.

111 As above.

112 As above.

113 Gillis and Speiss 2019 *The University of Chicago Law Review* 460.

114 BEUC 2020 6.

115 As above. Worth noting is that South Africa recognises the Treating Customers Fairly Principles and these principles are entrenched in the Policyholder Protection Rules 2018.

policyholder”.<sup>116</sup> The reason for this is that insurers are increasingly relying on big data analytics to better understand a policyholder’s individual price sensitivity as well as their likelihood to shop around or switch insurance at the point of renewal.<sup>117</sup> This is clearly not the correct way of premium pricing in insurance, as it does not focus on the risk of the policyholders but rather their individual propensity to switch insurance.<sup>118</sup>

These examples of unfair pricing practices must be circumvented and to do this it is proposed that there be a ban on unfair price optimisation practices when selling insurance products to consumers.<sup>119</sup> Insurers must set premium prices based on the individuals’ risk assessment and should not be setting prices based on consumers’ individual price sensitivity or their likelihood to switch insurance contracts.<sup>120</sup> In order to achieve fair outcomes in this regard, insurers should be required to publish information about the price differentials between their customers for transparency purposes. This could have the effect of increasing competitive pressures amongst insurers and ensure public and supervisory inspection to ensure that the pricing practices of insurers are fair towards consumers.<sup>121</sup>

### 5.5 Third example of discrimination: Data granularity

Data granularity refers to the level of detail considered in a decision-making process or represented in an analysis report, therefore, the higher the granularity, the deeper the level of detail.<sup>122</sup> Consequently, granular customisation may provide additional product availability and accessibility to policyholders,<sup>123</sup> but at the same time, some policyholders may find that they are not being offered specific types of cover at all, or that they are exposed to high non-risk related individual premiums and possibly discriminatory claims settlement decisions.<sup>124</sup> This may negatively impact the availability or even the affordability of insurance for certain parts of the population.<sup>125</sup> Further, this may result in diminished levels of consumer confidence in the insurance industry.<sup>126</sup> Insurance supervisors may also need to consider whether certain algorithmic practices may possibly undermine the underlying characteristic of risk pooling in insurance.<sup>127</sup>

---

116 BEUC 2020 6.

117 As above.

118 OECD 2020 13.

119 BEUC 2020 8.

120 As above.

121 As above.

122 IGI Global *What is Data Granularity* | IGI Global (igi-global.com) (last accessed 2022-04-16).

123 For example, it could benefit groups currently perceived to be of a higher risk, like elderly people, but who can demonstrate that their risk is lower.

124 BEUC 2020 8.

125 As above.

126 As above.

127 As above.

A core feature of private insurance policies is that customer segmentation occurs based on risk.<sup>128</sup> This means that higher-risk policyholders usually pay higher premiums compared to policyholders who are viewed as being lower-risk.<sup>129</sup> The proliferation of big data now allows insurers to consider a wider array of personal and behavioural data and charge corresponding premiums. Previously, risk-based pricing was based on a limited number of easily identifiable criteria.<sup>130</sup> This brings in new potential for discrimination as the use of AI may allow insurers to easily identify high-risk characteristics and result in categories of consumers no longer able to access or afford insurance coverage.<sup>131</sup>

Profiling the policyholder is increasingly used then it could reduce the availability, access, and affordability of insurance.<sup>132</sup> This type of risk segmentation will allow insurers to cherry-pick 'good risk' from 'bad risk' policyholders which will then ultimately lead to increased pricing differentiation between low- and high-risk consumers.<sup>133</sup>

Further to that, if personalised insurance products increase then there is serious concern that it may undo the traditional principle of "solidarity" or "risk pooling" that has always been at the core of insurance.<sup>134</sup> If insurers calculate every individual's personal risk and corresponding premium, insurance companies would no longer be spreading out the risk collectively between policyholders.<sup>135</sup> Hyper personalised risk assessments could leave certain individuals 'uninsurable' and lead to new forms of financial exclusion in the future.<sup>136</sup>

In Australia, for example, there is evidence emerging of more granular segmentation by insurers which is resulting in higher premiums and even potential financial exclusion of consumers.<sup>137</sup> The Competition and Consumer Commission found evidence that "[o]ver the past decade, insurers' methodologies for pricing insurance have become much more sophisticated and combined with access to better data, we have seen a shift towards more address-based risk assessment and pricing. As a result, insurance premiums are increasing, especially for those in high-risk areas".<sup>138</sup> One of the findings of the Inquiry was that "more granular

---

128 As above.

129 OECD 2020 8.

130 As above.

131 OECD 2020 8.

132 As above.

133 BEUC 2020 8.

134 Reinecke, van Niekerk and Nienaber (2013) 3 explain that the spreading of risk involves the transferring of risks of a community of exposed persons to a third party and for the risks then to be spread by the latter over that community; see also BEUC 2020 8-9.

135 BEUC 2020 9.

136 As above.

137 BEUC 2020 9.

138 As above; see also Australian Competition and Consumer Commission 'Northern Australia Insurance Inquiry' <https://www.accc.gov.au/system/files/Northern%20Australia%20Insurance%20Inquiry%20-%20First%20interim%20report%202018.PDF> (last accessed 2022-04-16).

pricing approaches, in particular address-based risk assessment, has been a key contributor to increased premiums for many policyholders.<sup>139</sup>

It will be necessary to have stronger oversight mechanisms in place when it comes to the use of personal data by insurers in terms of what they should be allowed to consider when selling insurance contracts and setting premiums.<sup>140</sup> Regulators will need to monitor whether the use of big data analytics has an impact on the insurability of high-risk consumers and may need to intervene to ensure people continue to have adequate access to insurance policies.<sup>141</sup>

The above discussion is evidence of how big data utilisation in insurance may result in varied discriminatory practices for policyholders. Furthermore, ethical considerations also need to be investigated as big data use will no doubt have certain ethical consequences.

## 6 Morality, ethics, and public policy in insurance contracts

### 6.1 Introductory thoughts

The issue of discriminatory practices in the insurance industry talks to risk classification and the fine line between differentiating between policyholders for sound economic reasons and attempting to punish or deal with concepts such as social structures within a country.<sup>142</sup> The question, as highlighted in the preceding paragraph, is whether such discriminatory practices are justifiable in such circumstances. Frees and Haung point out that the risk classification in insurance has been argued to be morally appropriate based on the actuarial sciences that underscore such differentiation.<sup>143</sup> In fact, the issue of morality becomes, in addition to the question of discrimination, a central feature in assessing the use of big data in the insurance industry. There are two sides to the issue of morality and ethics in the use of big data in the insurance industry. On the one hand, the use and processing of big data is to assess the insurability of the policyholder to determine whether the insurer is willing to provide insurance in the circumstances. This information links to, among other things, the moral risks of the policyholder which talks directly to the policyholder's character and integrity.<sup>144</sup> The use of big data in this context may result in discriminatory practices within the insurance industry, which was discussed in the preceding paragraphs.

---

139 As above.

140 BEUC 2020 9.

141 Limits to certain forms of granularity in risk-based pricing may need to be considered or limitations on the types of data points that are considered by insurance firms when setting insurance premiums.

142 Adapted from Frees and Huang 2021 *North American Actuarial Journal* 2.

143 Frees and Huang 2021 *North American Actuarial Journal* 2.

144 See, for example, *Potocnik v Mutual and Federal Insurance Co Ltd*.

The second aspect relates to the morality and ethics in the pre-contractual engagement with policyholders and the ultimate use of big data in contractual obligations.

Conduct that is legally permissive is not necessarily ethically permissive. Therefore, the role of big data may very well be regulated within the insurance industry through legislative and regulatory interventions, it does not necessarily consider the ethical considerations of the use of big data in the insurance industry. Hinman notes that “[e]very society has its set of moral rules or guidelines that establish the boundaries of acceptable behavior, that draw the line between good and evil”.<sup>145</sup> Hinman explains that these moral rules may be to protect people from harm or be related to issues of respect<sup>146</sup> and that the concept of ethics is the process of refining our moral beliefs.<sup>147</sup> The question then is what are those acceptable societal boundaries when considering the use of big data in insurance contracts?

## 6 2 Public policy in contracts

Central to the validity and enforcement of contracts generally is the concept of public policy and our views and understanding of public policy is, according to Southwood AJA, constantly evolving.<sup>148</sup> Like with public policy, big data is continuously developing and there may not be any easy answers to the questions surrounding the use of big data. Our courts have noted that in novel situations (specifically that of the duties of one of the contracting parties) one should turn to public policy for guidance.<sup>149</sup> Therefore, public policy may provide guidance both in assessing the validity of a contractual provision<sup>150</sup> and in establishing whether there are implied duties because of public policy considerations.<sup>151</sup> Yet, there is no clear guidance as of yet on whether public policy may be used to intervene where big data is used in determining whether a person may wish to conclude a contract.

Public policy has generally been considered to be part of the post-contractual phase e.g., determining whether a contractual obligation is valid and whether such an obligation would be enforced. The question is whether ethics and/or morality play a role in public policy considerations. Articulating exactly what public policy may be in a particular circumstance is often difficult.<sup>152</sup> To determine what public policy may

---

145 Hinman *Ethics: A Pluralistic Approach to Moral Theory* 5<sup>th</sup> ed (international edition) 4.

146 As above.

147 Hinman 5<sup>th</sup> ed 5.

148 *Price Waterhouse Coopers Inc and others v National Potato Co-operative Ltd* 2004 3 All SA 20 (SCA) para 24.

149 *Faiga v Body Corporate of Dumbarton Oaks and another* 1997 1 All SA 367 (W) 383.

150 See, for example, *Barkhuizen v Napier*.

151 *Faiga v Body Corporate of Dumbarton Oaks and another*.

152 See *Price Waterhouse Coopers Inc and others v National Potato Co-operative Ltd* para 24.

be, Joubert AJ notes that one should take into account “considerations of justice, equity, good faith, reasonableness, common sense and the like” in particular circumstances.<sup>153</sup> Public policy has also been found “that the concept of ubuntu and the necessity to do simple justice between individuals have been recognised as informing public policy in a contractual context”.<sup>154</sup> However, Sachs J (referencing Wessels) notes that public policy is linked to the “moral sense of the community”,<sup>155</sup> also called “public morality”.<sup>156</sup> More importantly, public policy is intrinsically linked to our Constitutional values.<sup>157</sup> Essentially, public policy forms the basis of the enforcement of contractual provisions, and may on occasion include the question of morality.<sup>158</sup> The issue of using big data in insurance contracts may require further assessment of whether ethics and morality actually play a role in insurance contracts.

### 6 3 Morality as an integral part of contracts

Ethics is generally considered to be a roadmap as to what is considered acceptable (or good) and what is considered to be unacceptable (bad) in society. Often the element of ethics is also linked to that of morality. Herein lies the challenge, as what is considered good or bad is often associated with the personal reflections and convictions of an individual and one would think that this has no place in the consideration of the law, especially something as objective as contractual engagements. Yet, our courts have recognised the possibility of morality in the entire life cycle of a contract, and this should be no different from insurance contracts.

The term ethics or morality is rarely used in the discussion of pre-contractual engagements. Rather, the terminology of “good faith” is used. Although the principle of good faith negotiations is acceptable and enforceable in South African law,<sup>159</sup> contracts of insurance are generally not negotiated. This does not conceptually preclude the requirement from the parties to “relate to each other in good faith”.<sup>160</sup> Furthermore, the information used to assess whether insurance will be provided to the policyholder is often associated with moral risk and hazards of the policyholder. It can then be said ethics and morality play a role in the pre-contractual engagements in the contract lifecycle.

---

153 *Faiga v Body Corporate of Dumbarton Oaks and another* 383.

154 *Everfresh Market Virginia (Pty) Ltd v Shoprite Checkers (Pty) Ltd* 2012 1 SA 256 (CC) para 61.

155 *Sasfin (Pty) Ltd v Beukes* 1989 1 All SA 347 (A) 350.

156 Du Plessis “Some thoughts on the consequences of illegal contracts” 2021 *Acta Juridica* 177.

157 *Price Waterhouse Coopers Inc and others v National Potato Co-operative Ltd* para 24.

158 See discussion in *Sasfin (Pty) Ltd v Beukes* 350.

159 See, for example, *Southernport Developments (Pty) Ltd v Transnet Ltd* 2005 2 All SA 16 (SCA); *Roazar CC v The Falls Supermarket CC* 2018 1 All SA 438 (SCA).

160 *Everfresh Market Virginia (Pty) Ltd v Shoprite Checkers (Pty) Ltd* 239.

Nothing seemingly prevents the contracting parties from agreeing, either expressly or impliedly to a moral and a legal duty in a contract.<sup>161</sup> However, it is necessary to note that duties cannot simply be implied into a contract by means of concepts such as ubuntu or good faith.<sup>162</sup> Another factor that is of importance in the insurance industry is the way big data is translated into standard-form contracts. Although the type of cover, the extent of the cover, and the policyholder may be individualised in an insurance contract, many insurance contracts are effectively standard form contracts that are used not only by one insurer but by the insurance industry as a whole. This may also be prejudicial to the policyholder. Ngcobo J in *Barkhuizen v Napier*, links morality to the philosophical foundation of contracts in noting that:<sup>163</sup>

In my view, to treat mass-produced script as sanctified legal Scripture is to perpetuate something hollow and to dishonour the moral and philosophical foundation of contract law. It certainly does not promote the spirit of openness central to our new constitutional order.

Our courts have also linked the enforcement of contracts to the concept of morality on more than one occasion. Ngcobo J, in the Constitutional Court case of *Barkhuizen v Napier*, also noted that:<sup>164</sup>

*Pacta sunt servanda* is a profoundly moral principle, on which the coherence of any society relies. It is also a universally recognised legal principle. But, the general rule that agreements must be honoured cannot apply to immoral agreements which violate public policy.

Herein, Ngcobo J links the enforcement of a contract on the basis of the sanctity of a contract to a moral principle but at the same time notes that sanctity of contracts cannot be applied to immoral agreements.<sup>165</sup> In these instances, contracts and that of their enforcement is linked to the morality of the matter, in other words, what is considered right or wrong in society in some way, shape or form. Although dealing with illegal contracts, the *par delictum* rule is worth mentioning as it will not allow for restitution of performance in an illegal contract wherein the contracting parties are equally morally guilty,<sup>166</sup> and a court will not enforce “a claim arising from a transaction which is base[d] in the sense that it violates morality”.<sup>167</sup>

Therefore, big data must not only be used in such a way as not to contravene anti-discriminatory legislation (as discussed in paragraph 3 above) but big data must also be used in an ethical and moral manner in all areas of the contract lifecycle.

---

161 See *Bark and Another, NNO v Boesch* 1959 2 SA 377 (T).

162 *Roazar CC v The Falls Supermarket CC* para 24.

163 *Barkhuizen v Napier* para 156.

164 *Barkhuizen v Napier* para 87.

165 See also *Juglal NO and Another v Shoprite Checkers (Pty) Ltd t/a Ok Franchise Division* 2004 5 SA 248 (SCA).

166 See, for example, the discussion in *Klokow v Sullivan* 2006 1 SA 259 (SCA).

167 *Shepstone v Shepstone* 1974 2 SA 462 (N) 465.

## 6 4 Ethical considerations in the use of big data

Big data and the ethical consequences raised through the use thereof is a complex issue. Ethical questions require cognitive work with a social dimension and these types of questions have in the past been performed by humans.<sup>168</sup> Big data analytics is the product of human choices, and they will affect human values.<sup>169</sup> Cultivating ethical sensibilities in big data is key in order to ensure that the use of big data is ethical and not unfair.<sup>170</sup> Therefore, designing the algorithms will require certain ethical considerations.

One important ethical consideration for the use of big data analytics in the insurance industry is the idea of “explainable AI” which encourages the notion that AI also needs to be able to respond to questions on “why” it has reached certain decisions.<sup>171</sup> This will encourage a certain amount of transparency in the decision-making process in order to understand why the algorithms have reached a certain decision. This can also enhance the fair treatment of consumers.

Notably, the OECD has put forward ethical considerations which should be considered by insurers utilising big data analytics. Some of the most important considerations raised by the institution include:

- human agency and oversight: meaning AI systems should empower human beings, allowing them to make informed decisions and fostering their fundamental rights;<sup>172</sup>
- technical robustness and safety: AI systems need to be resilient and secure, which means they need to be safe, ensuring a contingency plan is in place in the case of something going wrong, as well as being accurate and reliable;
- algorithms should ensure privacy and data governance: data protection and adequate data governance mechanisms must be guaranteed, taking into account the quality and integrity of the data, and ensuring legitimised access to data;
- transparency: the data, system, and AI business models should be transparent, there should be traceability mechanisms in place as well as the fact that AI systems and their decisions should be explained to their human counterparts.<sup>173</sup>

The OECD emphasises diversity, non-discrimination, and fairness as ethical considerations.<sup>174</sup> They state that unfair bias must be avoided, as it could have multiple negative implications, from the marginalisation of

---

168 OECD 2020 20.

169 Richards & King 2014 *Wake Forest Law Review* 426.

170 As above.

171 OECD 2020 26.

172 At the same time, proper oversight mechanisms need to be ensured.

173 OECD 2020 20.

174 As above.

vulnerable groups to the exacerbation of prejudice and discrimination.<sup>175</sup> It is thus clear that the potential for discrimination through the use of big data is an ethical issue that must be mitigated and managed by insurers and the developers of the algorithms.

A further consideration is the insurers' ability to repudiate the insurance contract based on incorrect information provided. In the past, the duty of disclosure rested on the policyholder as all the information relevant to the risk and for purposes of risk assessment is seen as being exclusively within the domain and knowledge of the policyholder and the insurer has no means to obtain such information itself. The underlying rationale in this regard is that the underwriter is deemed to know nothing and is solely dependent on the information supplied by the policyholder. This "traditional" approach presupposes that the policyholder is in possession of all material facts and information, and overlooks the emerging realities of big data which has the potential to place insurers in a more favourable position than their policyholders and which may result in informational asymmetry.

The use of big data and advanced analytics places insurers in a completely different position. Insurers are no longer limited to adopting a "passive" approach and can now place reliance on data obtained utilising technical capabilities or from third-party service providers such as social media. Insurers are able to obtain, mine, and analyse large sets of data for their own benefit and are thus able to gain an upper hand on their policyholders. This raises a second ethical consideration, the use of big data analytics in the pre-contractual stage and whether insurers would be able to repudiate a policy due to a non-disclosure if there was an error in the algorithm. Insurers would be required to develop certain checks and balances to ensure that the design of the algorithms is correct and that they are asking for the required information from policyholders. Insurers should only be allowed to repudiate a policy based on the current non-disclosure rules and should not be allowed to repudiate a policy based on inaccurate and incorrect algorithm designs.

Currently, there is no legislative or regulatory oversight of these ethical issues in the insurance industry in South Africa, and big data is left generally unregulated – exposing policyholders to risk and unreasonable contracting practices. Herein, we suggest that these ethical matters require consideration by regulators and insurers as the possibilities for discrimination in these areas are rife.

## 7 Conclusion

According to Frees and Huang, the "modern-day insurance industry is founded on the ability to differentiate, or discriminate, among risks, known as risk classification".<sup>176</sup> Risk classification is therefore at the

---

<sup>175</sup> As above.

<sup>176</sup> Frees and Huang 2021 *North American Actuarial Journal* 2.

heart of the industry. The Fourth Industrial Revolution has highlighted these practices as the insurance industry is becoming more reliant on big data, data analytics, and the use of algorithms in assessing its risks and issuing insurance contracts to policyholders.

Although big data provides great advantages to the insurance industry there is also a dark side to big data, which is often to the detriment of the policyholder. These elements include direct or indirect discriminatory conduct and result in the use of big data in the personalisation of offerings, personalisation of pricing, and data granularity. The South African legislative framework may mitigate some of the risks with legislative interventions that exist to protect the policyholder's privacy, which is found in section 9 of the Constitution prohibits unfair discrimination (including the prevention of harassment and hate speech), which right is given effect to in the Promotion of Equality and Prevention of Unfair Discrimination Act. This notwithstanding, the insurance industry has always differentiated between various types of people based on risk analysis and actuarial data, but this has generally not met the threshold of 'unfair discrimination'. Therefore, differentiation in the industry is something that already occurs and is generally accepted practice and such differentiation is not seen as 'unfair discrimination' due to the nature of insurance.

Big data is essentially a tool to assess the moral risks and hazards of the policyholder in order to determine the risk appetite of the insurer. At the same time, there are ethical considerations in the use of big data in the contract lifecycle which goes further than discriminatory consequences. It is therefore essential that insurers have appropriate systems in place to ensure that big data analytics do not generate discriminatory outcomes. If discriminatory biases in big data algorithms cannot be effectively avoided, then insurers should not install these systems.<sup>177</sup> Insurers should only use big data models if it can be clearly established with sufficient certainty that the model will not generate any prohibited discriminatory outcomes.<sup>178</sup>

The use of big data, data analytics, and algorithms has highlighted the potential of the above in risk classification in the insurance industry. Unless there are proactive interventions both legislatively but also within the insurance industry by means of codes of conduct and the like, it is not only possible but likely that these technological advances may deepen the social divide already ingrained within South African society. The uncircumspect and unregulated use of risk classification may ultimately be a barrier to social justice.<sup>179</sup>

---

177 BEUC 2020 13.

178 As above.

179 Frees and Huang 2021 *North American Actuarial Journal* 3.