Is 'risk' a useful concept to capture how AITs address uncertainty when designing and/or using AITs? In what ways might this concept be refined or elaborated and why?

1 Introduction

Risk is commonly used as a way to address questions of uncertainty at different stages of the design and use of AITs. In order to understand how this is possible we first need to understand what is meant by risk and uncertainty. Risk can be defined as 'the possibility of something bad happening'[1], however this has limitations when it comes to considering the different aspects which are involved. Uncertainty likewise is defined as 'something which is not known or something that is not known or certain'[2], this prediction can be made for things in the present as well as predictions of the future. With AITs there are many different aspects which need to be considered as to the uncertainties involved. For an AIT to make predictions of uncertainty there are two main approaches, with the relative uncertainty of an outcome or prediction being obtained by determining the probability of an output either from the error or from direct prediction using Bayesian networks. However, when considering AITs, questions of uncertainty are not always related to their direct predictions but instead these questions of uncertainty can relate to how the predictions are made as well as the situation in which the AITs are used.

To understand how useful a concept risk is in capturing how AITs address uncertainty we will start by considering how useful risk is as a concept in quantifying uncertainty, and how it's use in different domains affects this. After this we will consider some of the dilimitations which are associated with the concept of risk, and finally we will consider how the concept of risk can be used to address issues of uncertainty and how these the concept of risk is utilised for regulation. While considering these different aspects considerations will be made as to the limitations of risk as a concept in addressing uncertainty as well as how it can be refined.

2 Is 'risk' a useful concept to capture how AITs address uncertainty and how can it be refined?

Risk as a concept, while not able to address all aspects to do with uncertainty for AITs, can be used to address questions of uncertainty. This is possible as risk can be used to quantify uncertainties and the effects which they have. This is possible as the likelihoods of something happening or the accuracy of an output can be obtained from its uncertainty, allowing predictions of the risk to be made based on the severity of the potential outcome or the predictions. This can be seen in many different areas in which AITs are utilised, with the domains of use affecting how useful a concept risk is in capturing uncertainty. The use of risk as a concept to capture uncertainty and make decisions based on this is not new; it has been used in medicine and economics before the creation of AITs. This historic use of 'risk' to understand economic uncertainties has resulted in the normalisation of 'risk', and over time,

higher levels of risk becoming acceptable as things do not go wrong, resulting in companies becoming to 'big to fail'[3]. This lack of an understanding as to what different levels of 'risk' actually mean, shows that for risk to be a useful concept in addressing how AITs deal with uncertainty, an understanding needs to be had as to what risks are acceptable, and which ones are not.

However, in medicine, while the use of risk to capture uncertainty is nothing new, AITs often use risk as a way of making uncertainties relatable, as when making predictions as to the likelihood of someone having a disease, or probable outcomes such as complications after surgery[4, 5]. The use of AITs in medicine, however, illustrates two issues that occur when using risk to address issues of uncertainty. The first one is that risk does not give us information as to the context under which the uncertainty prediction was made; this has been seen by Caruana et al.[6] where an AIT was developed to assess the risk to someone that caught pneumonia, however the AIT failed to realise that people who had asthma were at high risk and so were admitted straight into the ICU and as such had a lower death rate. The second issue when using risk as a concept to address uncertainty is that while it may help doctors to prioritise certain things, knowing the risk gives no information as to how the risk can be reduced, as it generally does not give information as to the root cause[4, 5].

The use of AITs in the justice system has shown how well risk can address issues of uncertainty, as well as how using the predictions of uncertainty to determine risk can cause major problems. The use of the concept of 'risk' when making decisions in the justice system attempts to make these decisions more objective, as it tries to turn the uncertainty of something happening, such as the likelihood of re-offending pre-trial[7], into an actual objective quantity[8]. However, this attempt to use risk to make decisions more objective falls short, as in the case of pre-trial detentions, bias in both the data which informs the predictions of risk, as well as the inherent bias of the decisionmaker makes the decisions subjective[7, 8], with the interpretation of a given risk being individualistic[8]. The use of risk to address questions of uncertainty in pre-trial detention, as with medicine, does not consider the context in which the risk is determined. In the case of pre-trial detention risk is often considered so that it can be minimised, however this would mean that everyone would go into long term incarceration, which ends up increasing the risk of reoffending[9]. Therefore in the case of pre-trial detention, it would often be better to consider the concept of risk as only able to partly address the issue of uncertainty, as the concept of risk needs to balance multiple different components, such as the context and severity of the decision.

When thinking about how risk and how this addresses uncertainty in AITs, we need to also consider the uncertainties which can arise in the design and implementation phases. When developing an AIT there is always uncertainty associated in both how it works, as well as how we interact with it. To address these uncertainties, they can be thought about in terms of risks which need to be minimised[10]. However, the approach to deal with the uncertainties as to how AITs will work and the effects which they will have, can be done by considering the associated risks as ethical issues, which depending on the perspectives of the developers will be considered in different ways as the resulting AITs embody the designers' own standards as to what the acceptable levels of risk are[11]. These attitudes towards what are seen as acceptable levels of risk can change over time in society, as well as what constitutes a risk[11]. This is particularly relevant when considering how AITs address uncertainty in the environment in which they are applied, as they must be able to deal with the uncertainty of how they will be used. This is seen when considering social media, where the concept of risk can be used to capture the uncertainty of how people interact with it, as well as its effects on the individual.

In a society where we are obsessed with safety, the minimisation of risk is often prioritised[12, 13]. However, while using the concept of risk as a metric to capture how AITs address uncertainty may often make sense as 'risk' is often seen as a substitute for uncertainty, the

concept of risk covers more than just uncertainty, as has already been illustrated, making it a useful metric when making decisions, as it considers the impacts as well as their likelihoods of occurring. This ability for risk to consider the impact of certain events, allows for the consideration of black swan events (low probability high impact events) whereas if one was to just consider the uncertainty of an event in the design or use of the AIT, then these would be missed, as they often are unknown-unknowns[14, 15]. However such events can also be either missed when using risk to address uncertainties as they are not always negative, or they can be over prioritised missing more relevant issues[15].

While we have seen thus far that the concept of risk is able to address uncertainty when designing and using AITs, these abilities have limitations. When trying to use risk to capture the uncertainties involved risk is not always able to capture all aspects, especially as some elements which are seen as low risk but are highly uncertain could have high levels of significance to our understanding of the AIT. Further to this, what is considered a risk changes from person to person, this can prevent the use of risk in addressing uncertainty. However the ability for risk to include potential aspects which are not always considered, allows the concept of risk to be utilised to draw specific attention to things which are particularly relevant. While the use of the concept of risk may be able to address many of the uncertainties associated with AITs, it will by no means be able to address them all, as some of the uncertainties associated with AITs are related to their benefits and so never considered by the concept of risk, although these benefits may feedback in to how the overall risk of the system is viewed.

Recently the EU has proposed plans to use the concept of risk to regulate AITs[16, 17], likely due to its perceived ability to approximate the uncertainties involved as well there effects. However, while as has previously been seen the concept of risk can capture how AITs address uncertainty, the levels of risks which are associated with the different uncertainties vary with the individual[8]. These variations in the risks are which are associated with different uncertainties are due to the concept of risk being inherently subjective and biased, due to it enforcing a perception[8]. While these problems with the concept of risk may prevent its efficient use universally, as typified by the board range of different regulatory approaches around the world, it does allow the individual to address uncertainty from their perspective.

As the usefulness of the concept of risk to capture how AITs address uncertainty varies depending on the domain, the best way in which to refine the concept is to understand its limitations when applied to the specific domain. As while the concept of risk may be related to uncertainty, the notion of minimising the risk does not necessarily have the effect of reducing uncertainty as risk can not only be reduced by reducing the likelihood but also by reducing the severity[8]. It is also important to understand that knowing the risks this is not the same as knowing the dangers, and being able to avert them and other future uncertainties[14]. So as to refine the concept of risk it can be broken down in to two components; firstly as a heuristic as this attempts to make it a more objective when used for decision making and secondly as a manageable event where by it can be regulated against[18]. While risk may not be a universal concept[13] it is a useful concept to deal with as it allows for a greater understanding how AITs address uncertainty as if utilised well with the inclusion of both negative as well as positive effect it allows for the what if question to be asked[9].

3 Conclusion

The ‘risk’ is a useful concept to capture when considering how AITs address uncertainty, however as a concept it does have limitations which change depending on the domain of use which is being considered often due to what is emphasised. Through consideration of different domains we have been able to see here some of the issues which arise when using the concept of risk in considering how AITs address uncertainty. When considering economics

risk and uncertainty are seen as going hand in hand and often used interchangeable, however the ubiquity of risk in this context risks normalising it causing it to lose its meaning. However when considering medicine it was seen that while risk is important when addressing uncertainty, it is limited by the fact that it is unable to provide information as to how the risks can be mitigated. When considering the justice system the use of risk as a heuristic to address uncertainties which are derived from AITs is seen to be highly subjective taking on the individuals bias instead of becoming more objective as had been desired.

Thus it has been seen that when considering the concept of risk a greater understanding is needed as to what is meant when using the concept of risk to address uncertainty, as well as a more rigid definition what constitutes a risk especially as peoples understanding of risk has been seen to be subjective. But also the ability for risk to be used to consider aspects which are outside of just uncertainty allows the what if question to be asked however the concept of risk falls short as it only considers it form a negative perspective, but some uncertainties are beneficial.

References

1. Last accessed 10 May 2022. Available from: https://dictionary.cambridge.org/us/dictionary/english/risk
2. Last accessed 10 May 2022. Available from: https://dictionary.cambridge.org/dictionary/english/uncertainty
3. Langley P. Anticipating uncertainty, reviving risk? On the stress testing of finance in crisis. Economy and Society 2013; 42:51–73
4. Chen JH and Asch SM. Machine learning and prediction in medicine—beyond the peak of inflated expectations. The New England journal of medicine 2017; 376:2507
5. Begoli E, Bhattacharya T, and Kusnezov D. The need for uncertainty quantification in machine-assisted medical decision making. Nature Machine Intelligence 2019; 1:20–3
6. Caruana R, Lou Y, Gehrke J, Koch P, Sturm M, and Elhadad N. Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. *Proceedings of the 21th ACM SIGKDD international conference on knowledge discovery and data mining*. 2015 :1721–30
7. Esposito E. Beyond the promise of security: Uncertainty as resource. Telos 2015; 170:89–107
8. Green B. The false promise of risk assessments: epistemic reform and the limits of fairness. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*. 2020 :594–606
9. Green B and Chen Y. Algorithmic risk assessments can alter human decision-making processes in high-stakes government contexts. Proceedings of the ACM on Human-Computer Interaction 2021; 5:1–33
10. Johnson JG and Busemeyer JR. Decision making under risk and uncertainty. Wiley Interdisciplinary Reviews: Cognitive Science 2010; 1:736–49
11. Ananny M. Toward an ethics of algorithms: Convening, observation, probability, and timeliness. Science, Technology, & Human Values 2016; 41:93–117
12. Beck U, Lash S, and Wynne B. Risk society: Towards a new modernity. Vol. 17. sage, 1992

13. Starr C. Social benefit versus technological risk: what is our society willing to pay for safety? Science 1969; 165:1232–8

14. Amoore L. The politics of possibility. Duke University Press, 2013

15. Taleb NN. The black swan: The impact of the highly improbable. Vol. 2. Random house, 2007

16. Beverly AT. Decoding the Proposed European Union Artificial Intelligence Act. American Society of Intonational Law 2021 Sep; 25(20)

17. European Comission. Regulation of the European Parliament and of the Council - Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and amending Certain Union Legislative Acts. 2021; COM(2021) 206 final

18. Paul R. Risk: new issue or new tool in regulation and governance research? *Society, Regulation and Governance*. Edward Elgar Publishing, 2017