ARTIFICIAL INTELLIGENCE RISK AVERSION

Isariya Suttakulpiboon

ABSTRACT

I use survey to investigate people's risk attitude towards artificial intelligence (A.I.). If human and A.I. commit the same level of risk in different scenario, people still have a strong preference for human over A.I., given other things constant. I developed a tool called "accuracy premium" to measure the extra accuracy that A.I. must outperform human to make people feel indifferent. The result shows that the accuracy premium is significantly positive. Female, people with non-STEM background have significant and positive accuracy premia.

Keywords: Attitude Towards Risk, Artificial Intelligence

INTRODUCTION

My major motivation to write this paper arises from a casual conversation with a senior staff member in one of the major non-life insurance companies in Thailand. He is one of the leading team members developing and implementing an A.I.-based motor claim fraud detection system that would replace a current human-based system. He mentioned: “This is a very challenging task. We feel uncomfortable to fully implement A.I.-based system unless the accuracy is near perfect”.

“Why?” I asked.

“I don’t trust the machine. Even though our current system can poorly detect fraud and this A.I.-based system could bring a major improvement and significantly lower the operational costs, I still do not trust the machine”. He added, and I was puzzled.

How could the managerial team be reluctant to adopt the near perfect system that could potentially enhance the company’s competitiveness and significantly lower the overall costs? It has long been proven in many contexts, e.g. Amalberti and Deblon (1992) Diekmann (1992) Trippi and Turban (1992) and Wu et al (2014), that Artificial Intelligence (A.I.) could reduce model risks and could enhance performances, but what makes it hard to adopt? Is it risk aversion? Is it ambiguity aversion? Or we simply do not trust A.I. and still prefer human even though we are more prone to erroneous act?

The objective of this paper is to investigate “what does it take to make people adopt an A.I.” Most economists would use Willingness-to-Pay (WTP) to price the switching decision from human to A.I. WTP has played crucial role in many areas and is subject to numeral approaches. However, using WTP to assign a monetary value on the switching decision might be difficult to evaluate in the context of A.I. because it would be hard to imagine a proper monetary compensation of the switching cost. Therefore, using a non-monetary measure might be more appropriate. One interesting way to capture the net benefits of the switching decision is to ask, “how well A.I. need to perform better than human in order to switch from using human to using A.I.?” This measure could potentially be beneficial to the A.I. developer to create models that better predict outcomes without having to be 100% accurate; to the management team to make a better decision whether to adopt an A.I.; and to explore any cross-sectional differences among different groups of individuals on their ideas about A.I.

To achieve the objective, I use a survey to capture “*A.I. accuracy premium*" - a simple yet novel measure that capture extra accuracy that A.I. must outperform human to make people feel indifferent. Then, I try to discover any cross-sectional variations among different groups i.e. gender, background knowledge etc. to see which group might demand more accuracy premia. To answer the question “why we demand more accuracy premia”, I conduct a focus group with 10 selected respondents to gain detail insights, thus conclude the paper.

METHODOLOGY

I use questionnaire to measure the level of *A.I. accuracy premium* at the individual level. Consider the following scenario:

> You are at the hospital. You have been sick for a while and not sure if you have cancer or not (50-50 chance prior probability). The nurses have collected your blood sample, X-ray films, and run many tests already. Today you will be hearing the test result from one of the doctors: **the first doctor is a human and the second doctor is an A.I.** Both doctors will learn from your medical history and your current test result and they will only tell you whether you have cancer or not. No follow-ups. No small talk afterwards. Simply just a yes-or-no visit. **You also know that in the past, both doctors have diagnosed 100 patients and both of them have been misdiagnosed 10 patients (both have the same 10% likelihood of errors).**
>
> The first question is, if you can choose a doctor to hear the result, who would you choose, or do you feel indifferent?

The second question is, instead, you know that the A.I. doctors have misdiagnosed 8 patients whereas the human doctor have misdiagnosed 10. If you can choose a doctor to hear the result, who would you choose, or do you feel indifferent?

The third question is, instead, you know that the A.I. doctors have misdiagnosed 6 patients whereas the human doctor have misdiagnosed 10. If you can choose a doctor to hear the result, who would you choose, or do you feel indifferent?

And so on.

To construct the accuracy premia, each individual will be assigned a number from 0% to 10% For example, if the respondent answers "an A.I. doctor" to the first question, he or she will be assigned 0%. If the respondent answers "a human doctor" to the first question and "an A.I. doctor" or "I feel indifferent" to the second question, he or she will be assigned 2%, and so on.

I chose the following independent variables to explain variation in accuracy premium: FEMALE is a dummy variable equal 1 if the respondent is female. Previous literature, e.g. Eckl and Grossman (2008), Halek and Eisenhauer (2001), Jianakoplos and Bernasek (1998), Maxfield et al (2010) and Watson and Mcnaughton (2007), suggest the role of gender on risk attitude and risk perception. I expect that female respondents might demand accuracy premium from A.I. since they are relatively more risk averse than male respondents. STEM is a dummy variable equal 1 if the respondent obtained or currently enrolling in a STEM (science, technology, engineering and mathematics) or STEM-related degree. I expect that people associating with STEM-related degrees might have a strong preference of A.I. over human in general which might lead to a reduction or zero accuracy premium. Another important variable is A.I. is a dummy variable equal 1 if the respondent understands A.I. in general. I expect that if the respondents understand what A.I. is or how A.I. works, they might not require much accuracy premium. Other variables include INC1 and INC2 represent income level of the respondents where INC1 is a dummy variable representing respondents with monthly income range THB 15000 – THB 50000 (approx. $500 - $1667); INC2 is a dummy variable representing respondents with monthly income greater than THB 50000 (approx. $1667). EDU1, EDU2 and EDU3 represent respondents with bachelor's degree, master's degree or Ph.D. Age is an age level and EXP is a dummy variable equal 1 if the respondents have work experience greater than 5 years.

As suggested by Ferrari and Cribari-Neto (2004), I use beta regression model to explain a cross sectional variation in accuracy premium. Beta regression is one of the generalized linear model which allows dependent variable to be percentages. The density of the beta distribution is given by:

$$f(y;\mu,\phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma\big((1-\mu)\phi\big)} y^{\mu\phi-1}(1-y)^{(1-\mu)\phi-1}; 0 < y < 1$$

Let $y_1, y_2, .., y_n$ be independent random variables, where each y follows the above density with mean μ and unknown precision ϕ. The beta regression model is assuming that the mean of y can be written as:

$$g(\mu_t) = \sum_{i=1}^{k} x_{ti}\beta_i$$

Where each β_i is a regression parameter and x_{ti} are observations on k covariates which are assumed fixed and known. Finally, g( ) is a link function that maps [0,1] onto the real line. I use the logit link function to estimate the model.

In addition to using survey, I conducted a focus group with 5 selected respondents to gain insights as to why they might or might not require additional premia when using A.I. service.

DATA ANALYSIS AND RESULTS

The survey data is collected via Google Form from individuals in Bangkok metropolitan area using convenience sampling method. The total number of response is 606. Since the convenience sampling method is used, as shown in **Table 1**, the majority of the respondents are skewed towards female, younger respondents with STEM background, have middle income level, are studying undergraduate degree, no work experience and have little background in A.I.

The summary statistics in **Table 1** also show that the respondents demand, on average, 2.64% accuracy premia in order to switch from a human doctor to an A.I. doctor. **Figure 1** show preference shift towards A.I. once the difference between A.I. predictive accuracy is improved i.e. if there is no difference between an A.I. and a human doctor predictive accuracy, 55% of the respondents strongly prefer a human doctor to an A.I. doctor. However, the proportion reduces as the difference widens. It is worth noting that to switch from a human doctor to an A.I. doctor, most the respodents (almost 70%) only demand at most 2% accuracy premium while the rest might demand higher premia. Another interesting aspect is that even A.I. could perfectly predict the outcome, there is still a small number of respondents (around 2%) strongly prefer a human doctor or feel indifferent between a human doctor and an A.I. doctor.

Table 1: Summary Statistics (N = 606)

	Mean	Std.Dev.	Min	Median	Max
ACC PREM	0.0264	0.0252	0	0.02	0.1
FEMALE	0.6650	0.4731	0	1	1
A.I.	0.4778	0.5007	0	0	1
STEM	0.8275	0.3786	0	1	1
INC1	0.5891	0.3887	0	1	1
INC2	0.0788	0.2701	0	0	1
EDU1	0.3251	0.4695	0	0	1
EDU2	0.1280	0.3350	0	0	1
EDU3	0.0492	0.2169	0	0	1
AGE	24.374	7.1134	18	22	59
EXP	0.4532	0.4990	0	0	1

Figure 1: Respondents' Preferences over A Human Doctor vs An A.I. Doctor

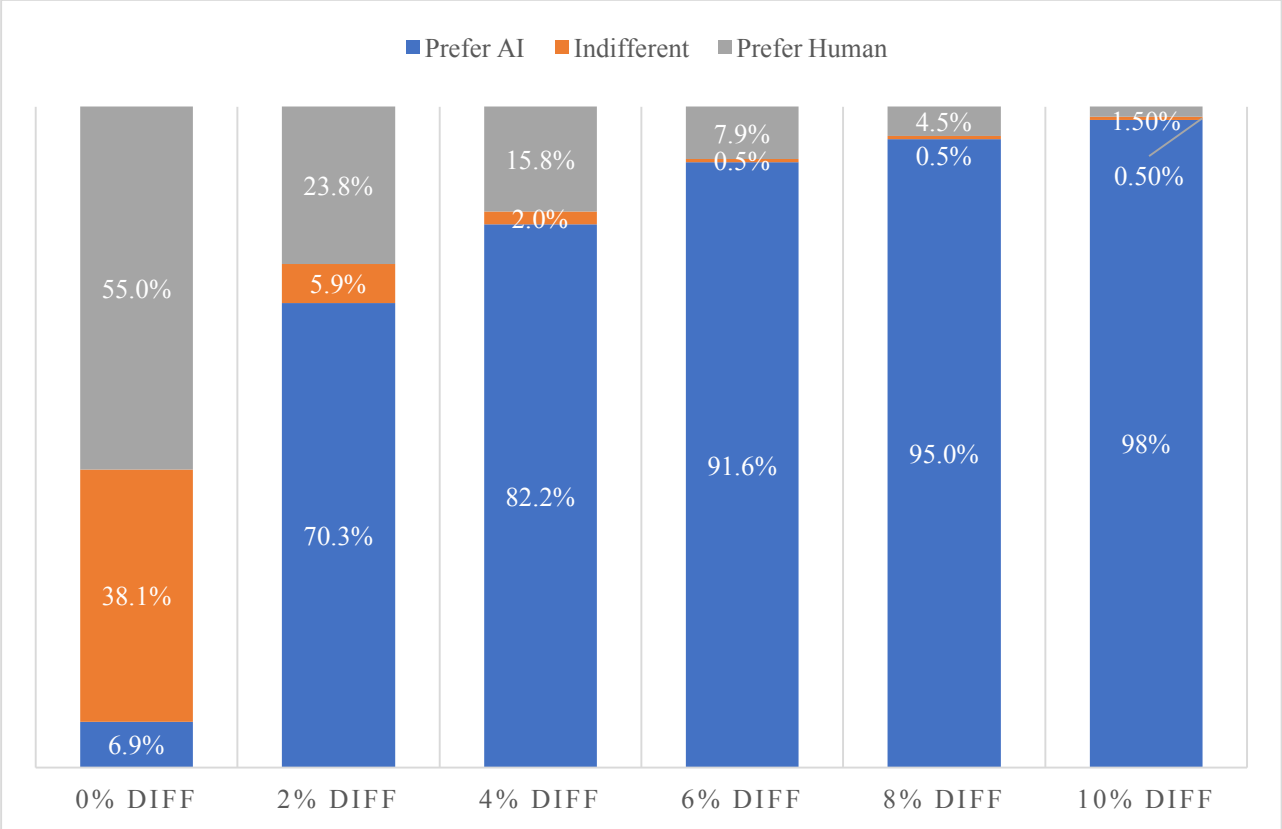

Table 2 reports the results from the beta regression model. The results suggest that female will generally demand *more* accuracy premium than male (p-value = 0.0165). Respondents with STEM-related background will demand *less* accuracy premium (p-value = 0.0698). I also found that respondents at the high-income brackets will demand *more* accuracy premium (p-value = 0.0134). Other variables remain insignificant.

Table 2: Results from Beta Regression

	Beta	Std.Err.	Z-Value	P-Value	Significance
INTERCEPT	**-3.445**	**0.480**	**-7.185**	**0.0000**	***
FEMALE	**0.356**	**0.145**	**2.398**	**0.0165**	**
STEM	**-0.031**	**0.017**	**-1.813**	**0.0698**	*
A.I.	0.119	0.146	0.815	0.4153	
INC1	-0.030	0.269	-0.113	0.9100	
INC2	**0.889**	**0.360**	**2.472**	**0.0134**	**
EDU1	0.415	0.285	1.457	0.1452	
EDU2	0.418	0.356	1.173	0.2408	
EDU3	0.642	0.453	1.418	0.1563	
AGE	-0.115	0.207	-0.556	0.5784	
EXP	0.101	0.302	0.335	0.7374	

Remarks: * 10% significant level, ** 5% significant level, *** 1% significant level

There are some interesting aspects to be discussed here. As explored in Eckel and Grossman (2008), Halek and Eisenhauer (2001), Jianakoplos and Bernasek (1998), Maxfield et al (2010) and Watson and Mcnaughton (2007), female is more risk averse and could potentially demand more accuracy premium. The more risk averse, the more certainty you prefer. The risk averse individual would prefer a less risky choice; therefore, demand a positive accuracy premium, and therefore, more than men do. Another potential explanation is that women do not easily trust the way men do, thus do not take excessive risks. According to the study by Fogel and Nehmad (2014), women tend to not voluntarily disclose their private information online (telephone number, address) because they simply do not trust other unknowns from online sources.

An insightful comment from one of the respondents is that A.I. could not be liable from the mistakes that they might commit: "I think the current legal system cannot punish A.I. the same way we punish human being. We know for sure we could bring the doctor to jail if there is a professional misconduct, but for A.I., I don't know… maybe we could punish the one who designed it, but it is much harder, to my knowledge, to do so". A.I. liability is poorly defined and may not be able to fully identify the potential ultimate damage or find the responsible liable person from such misconduct. As stated in Cerka et al (2015)[1], this could be one of the main reasons why we would demand an accuracy premium from A.I.

The significance of STEM variable suggests that respondents with STEM-related background feel more comfortable with A.I. and could accept the risk from A.I. prediction the same way they could do from the human prediction. One respondent agrees that "As I read and do more machine learning research, I know that A.I. could be one day as smart as human. We see chatbots and SIRI and other recommendation algorithms as used in Netflix or other websites getting smarter in knowing what we want and what we need. I would feel indifferent with A.I. telling me I have cancer if there is a same likelihood of error as the human doctor would commit". Another interesting aspect is that higher income individuals in my sample tend to demand excess accuracy premia.

1 Artificial Intelligence's ability to accumulate experience and learn from it, as well as its ability to act independently and make individual decisions, creates preconditions for damage. ... This means that with its actions AI may cause damage for one reason or another; and thus issues of compensation will have to be addressed in accordance with the existing legal provisions. The main issue is that neither national nor international law recognizes A.I. as a subject of law, which means that A.I. cannot be held personally liable for the damage it causes. In view of the foregoing, a question naturally arises: who is responsible for the damage caused by the actions of Artificial Intelligence?

SUMMARY

The objective of this paper is to measure the level of A.I. risk aversion and distrust – whether it can be compensated by an extra level of accuracy, or an accuracy premium. I found that the level of risk aversion and distrust exists and could be offset by an additional model accuracy. On average, my respondents agreed that an extra 2-3% accuracy on top of human-offered accuracy is sufficient; however, the percentages might vary cross-sectionally. Female is more risk averse and do not easily trust the A.I.; therefore, they might demand more accuracy premia. People with STEM-related background demand less accuracy premia due to their familiarity with A.I. and their strong preference of A.I. The results from this paper could be beneficial to modelers and A.I. developers who might want to build an A.I.-based system for people and companies. The accuracy does not have to reach perfection in order for people to use or for the company to adopt, but it has to offer an extra level of accuracy to compensate the level of risk aversion and distrust that people may have.

REFERENCES

Amalberti, R., & Deblon, F. (1992). Cognitive modelling of fighter aircraft process control: a step towards an intelligent on-board assistance system. International Journal of Man-Machine Studies, 36(5), 639-671.

Čerka, P., Grigienė, J., & Sirbikytė, G. (2015). Liability for damages caused by artificial intelligence. Computer Law & Security Review, 31(3), 376-389.

Diekmann, J. E. (1992). Risk analysis: lessons from artificial intelligence. International Journal of Project Management, 10(2), 75-80.

Eckel, C. C., & Grossman, P. J. (2008). *Men, women and risk aversion: Experimental evidence*. Handbook of experimental economics results, 1, 1061-1073.

Ferrari, S., & Cribari-Neto, F. (2004). Beta regression for modelling rates and proportions. Journal of Applied Statistics, 31(7), 799-815.

Fogel, J., & Nehmad, E. (2009). Internet social network communities: Risk taking, trust, and privacy concerns. Computers in human behavior, 25(1), 153-160.

Halek, M., & Eisenhauer, J. G. (2001). *Demography of risk aversion.* Journal of Risk and Insurance, 1-24.

Jianakoplos, N. A., & Bernasek, A. (1998). *Are women more risk averse?*. Economic inquiry, 36(4), 620-630.

Maxfield, S., Shapiro, M., Gupta, V., & Hass, S. (2010). *Gender and risk: women, risk taking and risk aversion*. Gender in Management: An International Journal, 25(7), 586-604.

Parente, S. L., & Prescott, E. C. (1994). Barriers to technology adoption and development. Journal of political Economy, 102(2), 298-321.

Trippi, R. R., & Turban, E. (1992). Neural networks in finance and investing: Using artificial intelligence to improve real world performance. McGraw-Hill, Inc..

Watson, J., & McNaughton, M. (2007). *Gender differences in risk aversion and expected retirement benefits*. Financial Analysts Journal, 63(4), 52-62.

Wu, D. D., Chen, S. H., & Olson, D. L. (2014). Business intelligence in risk management: Some recent progresses. Information Sciences, 256, 1-7.

Isariya Suttakulpiboon
Chulalongkorn Business School
Chulalongkorn University, Bangkok, Thailand, 10330
Email: isariya@cbs.chula.ac.th,