The Ethical Actuary: The industry's best defence against AI?

An independent research report by Louie Hart, 2026

Contents:

Preamble:

If there was a competition for number of AI-related videos watched and articles read over the past two years I don't think I would win. With that said, I feel confident that I would land somewhere in the top 1-2% globally and thus I have finally felt it time to put some pen to paper (In a metaphorical sense of course) in the form of an independent research report / opinions piece blend. I do this for a few reasons but chiefly, I hope that in transferring some of my own thoughts, in conjunction with the thoughts of many industry leading professionals, onto one document I might be able clear some headspace that has taken up much of my thought for the last 6 months. I also acknowledge that in the fast changing and accelerated-growth landscape of AI and Large Language Models the below could quickly become less relevant or accurate over time as sentiment, understanding, and of course the technology itself changes and shifts within our world. With that said, I'm an optimist and thought it necessary to provide something that actuaries (especially early career like myself) might hopefully be able to draw from going forward, even if that puts this piece at risk of aging not like a bottle of fine wine but rather a bottle of warm milk.

Introduction:

This report aims to explore the idea that ethical codes inherent to humans could be the modern actuary's best defence and differentiating factor against the coming, and in many ways ever-present, AI wave.
This comes at a time when I, and presumably most other budding actuaries in their early twenties, feel it not just necessary but vital to be thinking about how we can remain relevant and crucially: how we can continue to add legitimate value, in the foreseeably AI dominated corporate climate of the not too distant future.

I will tackle this topic by first unpacking why ethics plays a key, and indispensable, role within the Actuarial industry in a myriad of different ways. I will then explore the ethical behaviours of Artificial Intelligence and whether or not we can reasonably depend on AI, especially with growing autonomy, to act within the bounds of an ethics code that the actuarial industry, and humanity at large, deems acceptable.

Ethics in the Actuarial Profession:

Ethics is not merely a regulatory overlay within the realm of actuarial work, it is woven into the very fabric of daily practice. Being an Actuary means accepting and agreeing to certain ethical standards and codes of conduct laid out by the Actuaries Institute (Actuaries Institute, 2025). This code holds actuaries to a higher standard than that of most workplace codes of conduct, evidenced by principles like principle 5: speaking up – which not only demands that actuaries themselves behave and produce work in an ethical manner but also possess the moral courage raise concerns or issues that they identify in the work of colleagues around them (Actuaries Institute, 2025). This requires weighing personal risk against professional duty, navigating loyalty and confidentiality and having to make judgement calls when something has crossed the line.

Beyond this code of conduct, actuaries are faced with ethical considerations in many areas of their work and their ability to place these ethical considerations at the forefront of their practices and decisions is crucial not just to the actuarial industry but society at large. (Mark, *"The Importance of Ethics in Actuarial Decision-Making"*, 2025).

Ethical Risk Modelling

Risk modelling is to Actuary as football is to footballer - the word actuary is practically synonymous with risk modelling as this is a fundamental area of the actuarial realm. Risk modelling and risk assessments are hugely important to firms and as such so are actuaries (Wood, 2019), but with great power comes great responsibility. Actuaries are often faced with pressure from corporations for adjustments to risk assessments and models in order to meet profit expectations and goals (Mark, 2025). It is absolutely crucial that actuaries resist such pressures and present risk models that they truly believe to be true and fair reflections of both the relevant data and business climate. Remaining objective ensures a higher degree of long-term financial safety and reliability in spite of potential short-term profit motives (Actuarial Post, n.d.).

Actuarial Accuracy vs Fairness in Pricing

Another key area in which a very human element of fairness and ethics ought to be carefully considered and deployed by the modern actuary is in actuarial pricing models. When it comes to model training and development we are taught: "the more data, the *better* the model". This may be true in an absolute sense, however, the *best* models are not always

the best holistically. Briefly consider, for example, whether or not actuaries ought to be using data such as sensitive information, religion or ethnicity to price insurance premiums?

In Australia, for instance, Aboriginal and Torres Strait Islander (First Nations) people experience significantly higher mortality rates than non-indigenous Australians with a life expectancy gap of roughly 8-9 years (Australian Institute of Health and Welfare, 2024). Most people (and hopefully all actuaries) would agree that using someone's indigenous status as a risk-rating factor to price, for example, their life insurance policy would be morally wrong, unjust and ultimately - unethical.

In 2011 the European Court of Justice ruled, under EU directive 2004/113/EC, that insurance pricing discrimination on the basis of one's gender be prohibited creating 'gender equal insurance' (Schmeiser et al., 2014). From a purely mathematical perspective, the data was clear – women and men present statistically differentiated risk profiles and insurers argued that in passing this legislation economic efficiency would be lost (Insurance Europe, 2012). This law is not present in the Australian actuarial market which highlights the crucial aspect here – there is a high degree of subjectivity and human discretion required in the actuarial domain.

What works in one country might not work in another, people are seemingly okay with the use of their age for actuarial pricing but not their religion – all of these factors need to be carefully considered and evaluated with human ethics, empathy and experience when decisions which affect the lives and livelihoods of millions of people are at stake.

Can we trust AI to act ethically

Most of the dialogue surrounding ethics and artificial intelligence centres around the ethical *use* of artificial intelligence. Much could be written about this topic and how it relates to actuaries particularly as actuaries adopt the use of AI more and more into their everyday workflows. This paper, however, seeks not to evaluate the ethical use of artificial intelligence but instead the ability of AI to deploy its own set of ethical standards as it takes on increasing autonomy.

The Black Box Problem

Perhaps the biggest issue relating to the rapid development of artificial intelligence is that of the Black Box Problem. The essence of this problem is the fact that we are very often unable to understand or retrace how complex algorithms (like many AI models) arrive at their produced output, decision or answer (Elizabeth Louie, 2025). Not only does this make it impossible* for humans to examine or tweak any form ethical framework these systems may or may not possess but it draws into question whether these outputs can be trusted at all (Ali et al., 2023).

Moral Residue

If you are a human reading this report you are likely to have experienced that unsettling feeling, something of an internal signal, that occurs as a result of an outcome that technically benefitted you and yet to celebrate (even internally) feels misplaced. Wario memes would aptly describe this phenomenon by the phrase: "I've won but at what cost" but perhaps more formally, philosopher Bernard Williams called this feeling *Moral Residue*. Williams described the idea that even when one makes the 'correct' decision in a genuine ethical dilemma that something is still lost and that a morally serious person carries that with them (Gustavsson et al., 2025).

Now let us contrast that with AI. When people talk about an AI system being 'aligned' with human values and ethics they typically mean that the model has been trained, through techniques like reinforcement learning via human feedback, to produce a set of outputs that *look and feel* ethical (Zhang et al., 2022). These outputs, however, are merely the result of pattern matching and not conscience. The AI has no internal experiences with discomfort, nor does it lie awake at night contemplating the decisions (or outputs) that it made that day. The AI simply followed a decision tree (De Cremer & Narayanan, 2023). There is no inner conflict. No moral residue.

The Crux

Although it is reassuring to point out unique features of human morality that AI does not, seemingly by design, possess we ultimately require some stronger benchmark to evaluate AI systems as ethical moral agents. A Frontiers in Artificial Intelligence paper (2025) studied the responses of both ChatGPT and Claude to moral dilemmas and concluded that we cannot recognise AI systems as moral agents due to the fact that they

ultimately failed to meet the internal criterion of operating based on emotional mechanism (Barabadi et al., 2025).

Similarly, a fascinating structural argument is put forth by Elad Uzan (2025) who makes use of Gödel's Incompleteness Theorems to dissect and examine AI morality. The logic employed is that just like mathematics will always contain truths that lie beyond the world of formal proof, morality will always contain complexities that defy algorithmic resolution (Uzan, 2025).

Whether based on sampling studies or neatly dissected through a mathematics based philosophical lens, the findings and opinions of industry leaders and professionals seem, at least to me, to be resounding – artificial intelligence does not deploy its own set of moral frameworks and as such cannot be trusted to act ethically with autonomy.

Conclusion:

The above analysis has shown that humans have an ethical immune system that fires even when everything looks fine at a surface level. When considering the vast effect, and necessity, of ethics in the actuarial profession this could prove a huge advantage for young actuaries.

Consider, briefly, a scenario which draws on the earlier examination into the ethics surrounding actuarial pricing models. An actuary has built a pricing model – it's actuarially sound, commercially successful as well as being fully compliant with APRA* regulation. The firm is chuffed, regulators sign off and the model performs exactly as intended, apart from one (not so) small issue: premiums are landing disproportionately hard on a particular community – maybe low income earners, maybe people from a specific ethnicity. Nothing overtly illegal is happening. Nobody is complaining. Yet this actuary, our ethical actuary, can't sleep at night. That distinctly unsettling human feeling of moral residue - could be guilt, empathy, regret or perhaps a blend of some set of unnamed and infinitely complex human emotions – just might prompt them to do something, the right thing, for their fellow man. Something ethical. Something human.

*APRA stands for the Australian Prudential Regulatory Authority

References:

Actuaries Institute. (2025). *Code of conduct*. https://www.actuaries.asn.au/professional-standards-and-regulation/code-of-conduct

Mark. (2025, February 20). *The importance of ethics in actuarial decision-making*. Asia Pacific Actuarial Conference. https://aac2024.hk/the-importance-of-ethics-in-actuarial-decision-making/

Wood, R. (2019, March 28). *4 reasons why risk models are crucial for successful project management*. Safran. https://www.safran.com/blog/4-reasons-why-risk-models-are-crucial-for-successful-project-management

Actuarial Post. (n.d.). *8 risk hotspots that pose risks to quality actuarial work*. https://www.actuarialpost.co.uk/article/8-risk-hotspots-that-pose-risks-to-quality-actuarial-work-18360.htm

Insurance Europe. (2012, December 11). *Reaction to European gender ruling*. https://insuranceeurope.eu/news/2309/reaction-to-european-gender-ruling/

Schmeiser, H., Störmer, T., & Wagner, J. (2014). Unisex insurance pricing: Consumers' perception and market implications. *The Geneva Papers on Risk and Insurance – Issues and Practice*. https://doi.org/10.1057/gpp.2013.24

Australian Institute of Health and Welfare. (2024, July 2). *Health and wellbeing of First Nations people*. https://www.aihw.gov.au/reports/australias-health/indigenous-health-and-wellbeing

Ali, S., Abuhmed, T., El-Sappagh, S., Muhammad, K., Alonso-Moral, J. M., Confalonieri, R., Guidotti, R., Del Ser, J., Díaz-Rodríguez, N., & Herrera, F. (2023). Explainable Artificial Intelligence (XAI): What we know and what is left to attain Trustworthy Artificial Intelligence. *Information Fusion, 99*, Article 101805. https://doi.org/10.1016/j.inffus.2023.101805

Louie, E. (2025, February 23). *The black box problem*. Medium. https://medium.com/@humansforai/the-black-box-problem-c40d3c6f26fe

Blouin, L. (n.d.). *AI's mysterious 'black box' problem, explained*. University of Michigan-Dearborn. https://umdearborn.edu/news/ais-mysterious-black-box-problem-explained

Zhang, Z., Chen, Z., & Xu, L. (2022). Artificial intelligence and moral dilemmas: Perception of ethical decision-making in AI. *Journal of Experimental Social Psychology, 101*, Article 104348. https://doi.org/10.1016/j.jesp.2022.104348

De Cremer, D., & Narayanan, D. (2023). How AI tools can—and cannot—help organizations become more ethical. *Frontiers in Artificial Intelligence, 6*, Article 1093712. https://doi.org/10.3389/frai.2023.1093712

Barabadi, E., Fotuhabadi, Z., Arghavan, A., & Booth, J. R. (2025). Comparing AI and human moral reasoning: Context-sensitive patterns beyond utilitarian bias. *Frontiers in Artificial Intelligence, 8*, Article 1710410. https://doi.org/10.3389/frai.2025.1710410

Uzan, E. (2025, August 8). *What Gödel's incompleteness theorems say about AI morality*. Aeon. https://aeon.co/essays/what-godels-incompleteness-theorems-say-about-ai-morality