Ethical Dilemmas when using Models and Data

Ethical Conondrums with AI

Dr. Philippe J.S. De Brouwer

ALL RIGHTS RESERVED BY THE AUTHOR
prepared for: Model Risk Management for Financial Instititutions

October 26, 2026

Philippe De Brouwer

Academic and Professional Profile Dr. Philippe De Brouwer holds a background in theoretical physics and earned a second Master's degree in Business Engineering while working full time. His academic distinction includes solving the "fallacy of large numbers" puzzle, originally posed by P.A. Samuelson and unresolved for 38 years. In his Ph.D., he challenged the foundational assumptions of Harry Markowitz's Nobel Prize-winning Mean Variance Theory, developing the "Maslowian Portfolio Theory" and fundamentally advancing the field of investment advice.

Career Highlights Early in his career, De Brouwer transitioned from insurance to banking and from IT to asset management. At Fortis (now BNP Paribas), he played a key role in launching one of Belgium's first capital-guaranteed funds and was promoted to director in 2000. In 2002, he joined KBC, where he orchestrated the merger of four companies and became CEO of the consolidated entity in 2005. Under his leadership, the company climbed from 11th to 5th in the market, even amid heightened competition.

Following the financial crisis, De Brouwer led the creation of a new investment management company for KBC in Ireland, which quickly grew to manage around 1000 investment funds and approximately €32 billion in assets. In 2012, he broadened his expertise into financial risk management, focusing on statistics, analytics, data, and numerical methods.

Recent Roles and Publications De Brouwer served as Head of Analytics Development for the Royal Bank of Scotland Group and currently holds the position of SVP at HSBC in Kraków. There, he oversees the Model Review Centre of Excellence and serves as Global Head of Professional Practices for Model Risk Management. He also coordinates collaboration with universities.

He has authored numerous scientific papers and books, with his latest being "The Big R-Book: From Data Science to Big Data and Learning Machines." His work continues to influence both academic and practical approaches to finance, risk, and data science.

Wikipedia
Website

Contact
LinkedIn
e-mail
mobile: +48 790 715 002

DISCLAIMER / LEGAL NOTICE

1. Personal Capacity & Non-Affiliation

The content, presentation materials, oral remarks, analyses, and conclusions expressed herein are exclusively those of the author acting in his personal, independent capacity. They do not represent, reflect, or constitute the official views, corporate policies, academic positions, or commercial strategies of **HSBC**, **AGH University of Krakow (AGH)**, **Krakow University of Economics (UEK)**, or any of the author's other current or former affiliations, including their respective subsidiaries, parent entities, or affiliates.

2. Absence of Investment Advice or Recommendation

This material is prepared and distributed strictly for general informational and educational purposes only. Nothing contained in this presentation or handout constitutes financial, legal, tax, or investment advice. Specifically:

- It does not constitute an **investment recommendation** or **information recommending or suggesting an investment strategy** as defined under Regulation (EU) No 596/2014 of the European Parliament and of the Council (Market Abuse Regulation — MAR).
- It does not constitute **investment advice** or a **personal recommendation** within the meaning of Directive 2014/65/EU of the European Parliament and of the Council (MiFID II).
- It does not constitute an activity subject to regulatory oversight under the Polish Act of 29 July 2005 on Trading in Financial Instruments (*Ustawa z dnia 29 lipca 2005 r. o obrocie instrumentami finansowymi*).

3. No Offer or Solicitation

This presentation does not constitute an offer, invitation, solicitation, or marketing communication to buy, sell, subscribe for, or underwrite any financial instruments, securities, currencies, or underlying assets in any jurisdiction.

4. Disclaimer of Liability & Accuracy

While information and data presented have been obtained from sources believed to be reliable at the time of preparation, no representation or warranty, express or implied, is made regarding its accuracy, completeness, timeliness, or fitness for any particular purpose. The author, HSBC, AGH, UEK, and associated entities expressly disclaim any and all liability for any direct, indirect, incidental, or consequential losses or damages arising from reliance upon or use of this presentation or handout.

(c) Philippe De Brouwer

Contents

Contents 4

1 Defining Ethics 4

2 Defining Artificial Intelligence 7

3 Example: Insurance Onboarding 9

4 Further Ethical Conundrums 12

5 Conclusions 13

1 Defining Ethics

A Formal Definition – Cambridge Dictionary

Ethics

the study of what is morally right and wrong, or a set of beliefs about what is morally right and wrong.

Britannica: ethics, also called moral philosophy, the discipline concerned with what is morally good and bad and morally right and wrong. The term is also applied to any system or theory of moral values or principles.

How should we live? Shall we aim at happiness or at knowledge, virtue, or the creation of beautiful objects? If we choose happiness, will it be our own or the happiness of all? And what of the more particular questions that face us: is it right to be dishonest in a good cause? Can we justify living in opulence while elsewhere in the world people are starving? Is going to war justified in cases where it is likely that innocent people will be killed? Is it wrong to clone a human being or to destroy human embryos in medical research? What are our obligations, if any, to the generations of humans who will come after us and to the nonhuman animals with whom we share the planet?

Ethics deals with such questions at all levels. Its subject consists of the fundamental issues of practical decision making, and its major concerns include the nature of ultimate value and the standards by which human actions can be judged right or wrong.

The terms ethics and morality are closely related. It is now common to refer to ethical judgments or to ethical principles where it once would have been more accurate to speak of moral judgments or moral principles. These applications are an extension of the meaning of ethics. In earlier usage, the term referred not

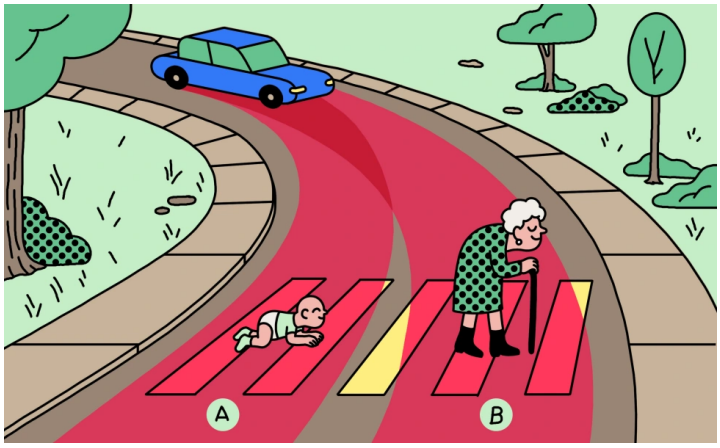


Figure 1: *You can only correct the course of the car a little, so you can reduce the deaths from two to one. Do you choose track A or B? — source: MIT Technology Review — Hao, 2018*

to morality itself but to the field of study, or branch of inquiry, that has morality as its subject matter. In this sense, ethics is equivalent to moral philosophy.

Although ethics has always been viewed as a branch of philosophy, its all-embracing practical nature links it with many other areas of study, including anthropology, biology, economics, history, politics, sociology, and theology. Yet, ethics remains distinct from such disciplines because it is not a matter of factual knowledge in the way that the sciences and other branches of inquiry are. Rather, it has to do with determining the nature of normative theories and applying these sets of principles to practical moral problems.

Ethical dilemmas are more important than ever. With the next figures we explore a few variations of the trolley problem. The Trolley Problem is no longer just a thought experiment for philosophy students; it is a technical requirement for autonomous systems. Engineers must decide how a car should value a child's life versus an elderly person's life in a situation where a collision is unavoidable.

AI Dilemma: Kill/Save the Baby or Old Person?

It appears that the choices people make is somehow determined by culture. In Western culture a majority of people will choose to save the baby, in West Asian cutltures typically the older person is saved.

The AI Version of the Trolley

(c) Philippe De Brouwer

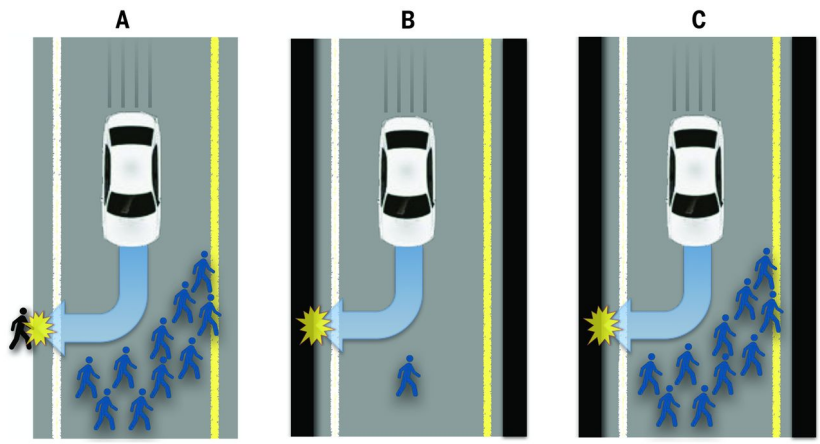


Figure 2: *source: Science — Bonnefon, Shariff, and Rahwan, 2016*
Should we veer and kill the person who is on the pavement? What if there is only one person on the street and there is a wall? How many people to change choices?

Is it ethical to remove the house of a squirrel to build your house?

Figure 3: *Is removing a tree that houses a squirrel ethically correct when building your home? What if an advanced society wants to make Earth their home?*

Most people will find that removing the tree that houses the squirrel is morally acceptable. If the squirrel would be another human, typically (in mod-

ern times) we would not accept this. However, we and the squirrel share a common ancestor that lived ca. 180 million years ago.

We are only separated by 180 million years, and we do not see the squirrel as a valid discussion partner for the location of our house. Now, imagine an alien society that has a 180 million years earlier start of their civilisation than us. They will probably not recognise us as intelligent partners, and just like the squirrel we will be defenceless against their technology if they want to take our planet.

Suddenly the same problem feels very different.

What is ethically acceptable depends on your viewpoint

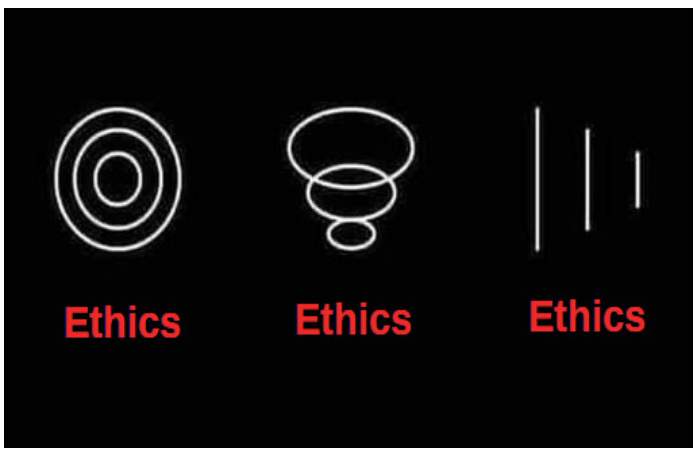


Figure 4: *Ethics is complex and it depends on your point of reference and you frame.*

2 Defining Artificial Intelligence

A Definition of AI

Artificial Intelligence

A cross-disciplinary approach to understanding, modelling, and replicating intelligence and cognitive processes by invoking various computational, mathematical, logical, mechanical, and even biological principles and devices. —Brundage, 2015

or ...

AI in Britannica.com

the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings

or ...

Defining AI Leaves Simple Questions Open

- What is exactly Intelligence?

  the ability to acquire and apply knowledge and skills (Oxford Dictionary)

- How to Measure Intelligence?

  IQ? EQ? What about logical, interpersonal, spacial, musical and many other forms?)

- What is Self-consciousness? Sentience? Life?

  ?

- When should an AI get animal or human rights?

  ?

Intelligence is defined by the Oxford English Dictionary as the ability to acquire and apply knowledge and skills. However, quantifying this capability remains a complex endeavor. Traditional psychometric assessments, such as the Intelligence Quotient (IQ), focus primarily on logical-mathematical and spatial reasoning. In contrast, modern frameworks account for broader dimensions of cognition, including Emotional Intelligence (EQ) alongside interpersonal, musical, bodily-kinesthetic, and intrapersonal capacities. Measuring intelligence therefore requires recognizing that mental capability is multifaceted rather than a unified, linear metric.

Beyond cognitive processing lie the deeper ontological concepts of sentience, self-consciousness, and life itself. Sentience refers to the capacity to experience subjective feelings and sensations, whereas self-consciousness represents an explicit awareness of one's own existence and internal mental states. Defining the precise moment when physical or synthetic organization transforms into a living, conscious entity is one of the most formidable questions in contemporary philosophy. As non-biological systems demonstrate increasingly sophisticated adaptive behaviors, the boundary between automated execution and genuine self-awareness becomes progressively obscured.

This philosophical ambiguity directly influences the discourse surrounding moral status and legal rights. Society must inevitably address whether moral consideration ought to depend upon an entity's physical origin or its subjective

internal experience. Consider the varied modes of artificial creation depicted in modern fiction: an entity constructed through mechanical and digital engineering, such as the holographic companion Joi in *Blade Runner 2049*; a synthetic being crafted through advanced biological engineering, such as Rachael in the *Blade Runner* universe; or an inanimate object brought to life through magic, as seen in the film *Ted*. Regardless of whether an entity comes into being via biological synthesis, computer engineering, or supernatural phenomena, humanity will need to confront a critical question: if a created being demonstrates verifiable sentience and self-consciousness, on what ethical ground can basic human or animal rights be withheld? Ultimately, moral standing may need to be determined not by how an entity was constructed, but by its capacity to feel, reflect, and experience existence.

3 Example: Insurance Onboarding

Car Insurance Claims Data: Acceptable Data

The data contains

- driving related:
  - average daily travel
  - professional or private use
  - living in urban or city area
- car related:
  - type of car (eg. minivan, sports car, sedan, etc.)
  - age of the car
- behaviour on our books:
  - time with us
  - old claims
- driver related:
  - has driver's licence been revoked
  - penalty points on driver's licence
  - years on job (work related)

Car Insurance Claims Data: Protected Features

. . . but also contains

- related to gender and family:
  - number of kids at home
  - single parent or not
  - marital status
  - gender

(c) Philippe De Brouwer

- education and profession
  - education
  - occupation
- wealth related:
  - value of the car
  - value of the home

Protected features are variables that, for ethical or legal reasons, should not influence a decision-making model. In insurance, using factors like gender or marital status can lead to discriminatory pricing, even if those factors are statistically predictive of risk.

The cost of being politically correct?

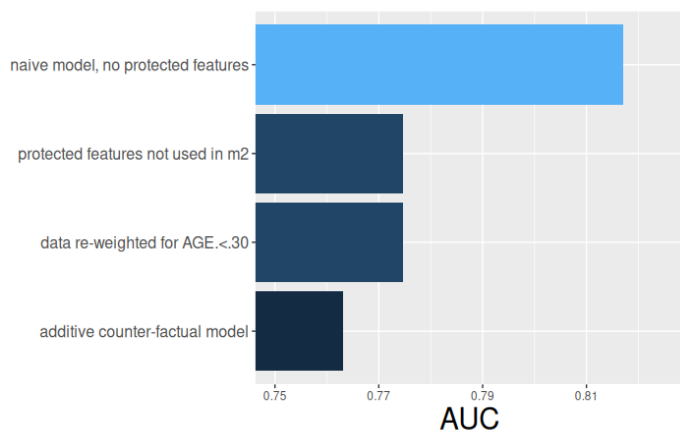


Figure 5: *The discriminating power of different models. The higher the AUC, the stronger the model, the lower the losses on the portfolio, the higher profit (or lower prices) can be.*

Area Under the Curve (AUC) is a foundational statistical metric used to evaluate the performance of binary classification models, particularly the Receiver Operating Characteristic (ROC) curve in risk analysis. Within the context of insurance underwriting, AUC measures a model's discriminative power—its structural capability to correctly rank-order policyholders so that individuals who will eventually file an accident claim receive higher risk scores than those who will remain claim-free.

- **Discriminative Power:** AUC quantifies the probability that a randomly selected high-risk customer is assigned a higher risk score than a randomly chosen low-risk driver. An AUC score of 0.5 denotes performance no better than random guessing, whereas a score of 1.0 indicates a theoretical, flawless predictor.

- **Portfolio Efficiency:** Stronger models with higher AUC scores allow insurers to fine-tune pricing tiers, minimize cross-subsidization, reduce portfolio-wide losses, and offer more competitive pricing to safe drivers.

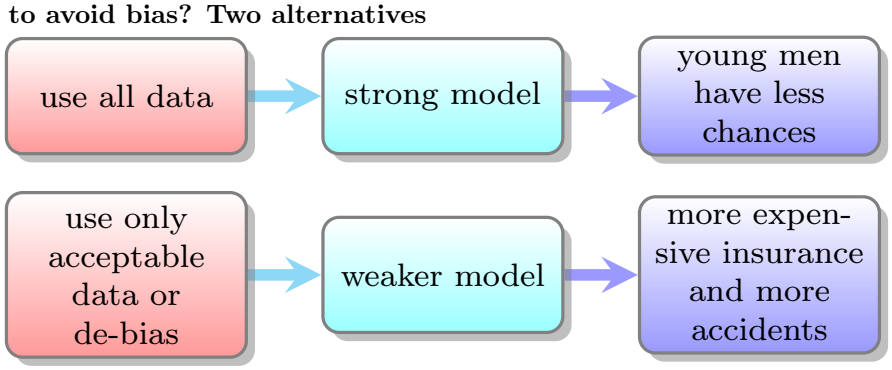
- **The Ranking Limitation:** AUC evaluates *relative ordering* rather than absolute risk probability or financial magnitude. It confirms whether high-risk drivers are positioned above low-risk drivers, but it remains entirely blind to the actual monetary value of potential claims.

Because AUC is strictly a rank-order metric, it cannot be directly converted into a precise monetary figure or a specific financial impact. When compliance frameworks or ethical mandates require data scientists to strip out protected features—such as gender, marital status, or socioeconomic indicators—the model's AUC inevitably drops, representing a loss in statistical precision.

However, translating this drop in AUC into an exact dollar amount or premium hike for an individual policyholder is fundamentally impossible through the metric alone. A reduction in AUC from 0.85 to 0.80 indicates degraded portfolio discrimination, but it does not dictate how pricing tables must shift.

Quantifying the financial cost of eliminating bias requires mapping the altered risk distributions against commercial underwriting rules, loss cost projections, and regulatory pricing constraints. While a weaker model inherently introduces portfolio inefficiencies, the exact financial burden experienced by specific consumer groups remains concealed behind complex actuarial formulas rather than visible within the AUC score itself.

How to avoid bias? Two alternatives



(c) Philippe De Brouwer

4 Further Ethical Conundrums

AI Governance & Ethical Horizons

Data Governance: Autonomy, Security, and Market Power

- **Informed Consent & Behavioral Autonomy:** Preserving genuine individual free will by combatting deceptive dark patterns, pervasive surveillance, and algorithmically driven behavioral manipulation.
- **Data Safeguarding & Regulatory Compliance:** Establishing robust cyber-security protocols to prevent systemic breaches while ensuring strict adherence to global privacy standards (e.g., GDPR, CCPA).
- **Intellectual Property & Anti-Competition:** Clarifying ownership rights over training corpuses, user prompts, and synthetic outputs to mitigate platform lock-in and data monopolies.

Algorithmic Governance: The FATE Framework

- **Systemic Fairness & Data Curation:** Mitigating historical bias in automated scoring while addressing the ethical dilemma of active data intervention versus objective representation.
- **Legal & Moral Accountability:** Defining clear chains of responsibility and legal liability when autonomous systems commit errors or cause tangible harm.
- **Transparency & Explainability:** Requiring open disclosure of algorithmic deployment alongside auditable, human-interpretable models for high-stakes decision-making.

Societal Horizons: Alignment, Moral Status, and Agency

- **The Alignment Problem & Lethal Agency:** Ensuring superintelligent goals remain compatible with human survival, while setting absolute ethical boundaries on lethal autonomous weapons systems.
- **Post-Human Moral Status:** Determining the legal rights, digital bodily autonomy, and shutdown protections for artificial sentience, cybernetically enhanced humans, and uploaded consciousness.
- **Normative Authority ("Who Decides?"):** Identifying which global bodies hold the legitimate mandate to construct, regulate, and enforce universal standards of machine ethics.

5 Conclusions

Visual Bias 1

Visual Bias 2

Fairness is a matter of perspective

We do have an innate sense for what is fair and what not, however that sense is heavily biased towards ourselves. For example, we don't see it as a moral issue to build a house and destroy the home of a squirrel in the process. The squirrel, however, is only separated from us by 180 million years of separate evolution. From a society that is a million years ahead of it might seem obvious to simply make earth their home and get rid of us.

If that seems unfair, turn it around. Would it be fair to colonise a planet inhabited by squirrel-like mammals and make it our home?

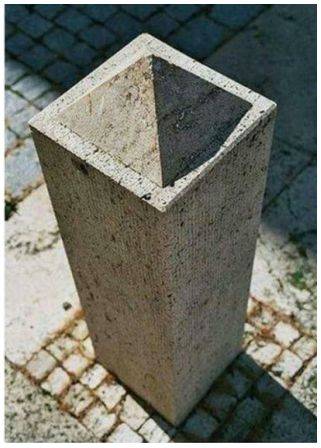


Figure 6: *Visual biases as clear: is the pole hollow or is there a pyramid on top of it?*

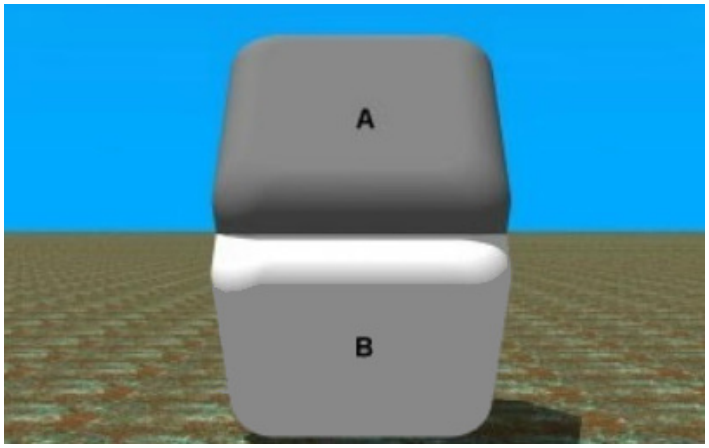


Figure 7: *Visual biases cannot be avoided. We know that A and B have the same shade of grey, but we cannot see it correctly.*

Figure 8: *Fairness is a matter of perspective*

Conclusions

Intuition	Innate **ethical intuitions** are insufficient for AI.
Perspective	**Subjectivity**: Ethical stance depends on perspective.
Judgment	**Context**: Morality requires context-sensitive judgment.
Trade-offs	**Balance**: Ethical decision-making involves trade-offs
Responsibility	**You have an individual and non-transferable responsibility.**

References

Bonnefon, Jean-François, Azim Shariff, and Iyad Rahwan (2016). 'The social dilemma of autonomous vehicles'. In: *Science* 352.6293, pp. 1573–1576.

Brundage, Miles (2015). 'Taking superintelligence seriously: Superintelligence: Paths, dangers, strategies by Nick Bostrom (Oxford University Press, 2014)'. In: *Futures* 72, pp. 32–35.

Hao, Karen (2018). 'Should a self-driving car kill the baby or the grandma? Depends on where you're from'. In: *URl:https://www.technologyreview.com/s/612341/a-global-ethics-study-aims-to-help-ai-solve-the-self-drivingtrolley-problem/*.