MORAL AND ETHICAL LEADERSHIP IN A DIGITALIZED FINANCIAL INDUSTRY

CAPSTONE PROJECT

TIAS Executive MBA

Ing. Jop Versteegt CPC

29 September 2020

Primary supervisor: Prof dr. Lans Bovenberg

Secondary supervisor: Dr. Eric Dooms

How do we make our moral and ethical decisions?

This seemingly simple question is difficult to answer—especially if a robot will be answering it for us. Robotization and digitalization are common phenomena in our daily lives. In the near future, robots will be making more and more decisions for us. This raises questions concerning our future with them, as well as with their more intelligent successors. Will we be able to control artificial-intelligence (AI) robots, including their moral and ethical decision-making processes? To do this, we must first understand how we humans make our own moral and ethical decisions.

In the financial industry, we have also begun to use AI Robots, which are able to analyze data efficiently and are helpful in reducing risks, analyzing future trends, and even preventing damages and detecting fraud. They are helping us to be more proactive in client services. These AI Robots are already enabling us to process greater volumes of data, in greater personal detail, in order to perfect our responses to the needs of our clients.

The use of AI Robots undoubtedly provides more power. It is important to remember, however, that greater power comes with greater responsibility. Responsible leadership should focus on achieving long-term goals to the benefit of our natural environment and society for tomorrow and beyond. Given the vast replication skills of AI robots, responsible leadership of AI Robots will require more explicit and ethical decision-making, in order to ensure that the robots act within set boundaries to achieve these long-term goals.

This Capstone project provides new insight on moral and ethical leadership in a digitalized financial industry. I hope that reading about the process of the research, the interview results, and the final conclusions will form an inspiring journey.

I am convinced that this research is only an initial stepping-stone, and further steps should follow.

CONTENT

The challenge posed by the lack of a decision framework

The financial industry is a complex industry, in which trust is the most crucial factor. It is also a money-driven environment, characterized by constant pressure on ethics and morality. Public opinion has pushed financial institutions to shift their focus from short-term profit to long-term sustainable business. Moral and ethical decisions are increasingly being challenged with regard to their long-term perspective. Robotization and digitalization are common practices in the financial industry, and AI is already being implemented in investment robots, client services and many other aspects. it is important to consider what we have (or could have) learned from these initial implementations. Have they been focused on providing better client services or reducing risks, or have they instead been aimed at increasing financial benefits to the company? What kind of leadership could be capable of fulfilling the balanced ethical and moral decisions in the industry, thereby ensuring that self-interest does not result in biased decisions? How can we supervise such leadership?

What strikes me is the lack of a generally accepted decision framework[1] on which ethical and moral decision-making in the financial industry can be based.[2] In the absence of such a framework, how can we supervise those who are facing actual challenges within their environments, and how can we ensure that they will make moral and ethical decisions in the best interest of their clients? From a more future-oriented perspective, what will be the impact of implementing new AI robot technology in the financial services sector? How can we make moral and ethical decisions while robots are selecting and providing advice on financially impactful products?

Although future moral and ethical questions might not be very different from those that are currently at play, the potential impact of robotic decision-making and replication skills obliges us to make to make such questions more explicit before having them implemented by AI robots. The goal of this Capstone project is to help leaders make the best possible ethical and moral decisions by harnessing the substantial powers of AI robots to the benefit of all of us.

"We need the as yet undiscovered fences from an undiscovered country.
But at least, we do know what a fence is." (Gillies, 2019, p. 14)

Given the challenges facing the financial industry, and in light of future developments in the field of AI, it is reasonable to question whether it would be better to adapt an existing framework or to develop a new one on which to base moral and ethical decisions. Furthermore, we should question how such an adapted or new framework is likely to impact human leadership and its governance. Moreover, we must determine how we will monitor, develop, and control the morality and ethics of AI robots that will be making such decisions in the nearby future, based on our input.

Example from outside the financial industry

In developing self-driving cars, Mercedes made the moral and ethical decision that the driver should be protected at all times (Sorrel, 2016). This Capstone project focuses less on the outcome of such decisions than on the manner in which such decisions are made. It further explores the questions that decision-makers discuss among themselves before arriving at their final statements. In addition, we examine what was required of leaders in order to make such decisions and the challenges posed by their governance model.

---

[1] *In this project, a framework is understood as a structure within which to develop a systematic approach for making decisions on moral and ethical questions (e.g., the "ethics check questions" presented by Norman Vincent Peale and Kenneth Blanchard; (Peale & Blanchard, 1988).*
[2] *The Banker's Oath requires employees to perform their jobs with integrity and care, in addition to carefully weighing the interests of all parties involved. The oath nevertheless pays no attention to the manner in which moral and ethical questions should be answered.*

The main research question of this Capstone project is as follows: *What should be the moral and ethical decision framework in a digitalized financial industry?*

Before answering this research question, it is important to provide more context by answering the following more specific questions:

- What are morality and ethics?
- What is a moral and ethical decision framework?
- How will morality and ethics in the financial industry develop in the future?

1. What are morality and ethics?

What is the definition of ethics and morality? The word *ethics* is derived from the Greek *ethos*. The word morality comes from the Latin *mos*. Sharing a similar etymological background, ethics and morality are often used in combination with each other. The ethical or moral sciences aim to define whether specific actions/behaviors can be defined as good or bad. In this paper, we combine the two terms to discuss ethical and moral behavior. This usage reflects on behavior, while also referring to specific actions. Both of these aspects are addressed in this project.

Morality

"Morality is the set of views, decisions and actions in which people express what they think is good or decent" *(Skeepies, 2011, p. 1).*

Ethics

"Ethics are the systematic reflection on morality."

"Ethics, the moral principles, values and rules that govern our decisions and actions with respect to what is right and wrong, good and bad" *(D.A. Buchanan, 2017, p. 61).*

Individual and business ethics
When discussing morality and ethics, it is important to distinguish between individual choices and those made from an organizational perspective. Johnsen, Scholes, and Whittington developed a three-level framework (G. Johnson, 2008).

Level 1: *Individual ethics*

Decisions and actions of individual managers and their ethical principles.

Level 2: *The organization's ethical stance*

The minimum obligations to stakeholders and society at large that are to be exceeded

Level 3: *Corporate social responsibility*

The ethical stance in practice, taking into account the interests of various stakeholders

This paper focuses primarily on Levels 1 and 2, with an emphasis on individual behavior, although it also addresses ways in which morality and ethics can be incorporated into corporate behavior. The perspective is oriented toward the future, exploring the impact of AI robot decision-making.

2. What is a moral and ethical decision framework?

A moral and ethical decision framework is understood to be a structure within which to develop a systematic approach on how to make decisions on moral and ethical questions. The focus of such a framework is not on the outcome of the answer or on developing a moral compass or moral values. Instead, the paper focuses on the process through which decisions have been reached or how they can be made in such way that all relevant aspects are included.

One example of a moral and ethical framework was developed by Norman Vincent Peale and Kenneth Blanchard in 1988, based on their "ethics check questions" (Peale & Blanchard, 1988).

- Level 1: Is it legal?
  Will I be violating either civil law or company policy?

- Level 2: Is it balanced?
  Is it fair to all concerned in the short term as well as the long term? Does it promote win-win relationships?

- Level 3: How will it make me feel about myself?
  Will it make me proud?
  Would I feel good if my decision was published in the newspaper?
  Would I feel good if my family knew about it?

3. How will morality and ethics in the financial industry develop in the future?

In *The Wealth of Nations* (Smith, The Wealth of Nations, 1776), Smith argues that our individual need to fulfil our own self-interest results in societal benefit. Known as the "invisible hand," this phenomenon combines with the division of labor in an economy to generate a web of mutual interdependencies that promote stability and prosperity through the market mechanism.

The ethics of the financial industry have been influenced by several incidents in which organizations focused on greed and bonuses have wasted their opportunities. As demonstrated by the financial crisis of 2008, self-interest (an active contributor to the crisis) does not necessarily promote social stability.

The industry is now focused on changing its morality, and we sincerely hope that its ethical and moral behavior will reflect true evidence of such change in the near future. The focus should be on the long-term goals, as also presented in Simon Sinek's *The Infinite Game* (Sinek, 2019).

According to a common contemporary belief, a focus on ethics and moral behavior is needed in order to increase trust in financial institutions. Recent examples in the media include the ING/Ralph Hamers case (Simons, 2016) (Peters T. , 2018) and, earlier, Joris Luyendijk's *This Cannot Be True* (Luyendijk, 2015). These examples demonstrate that, according to common opinion, the financial industry still has a long way to go.

A moral and ethical framework could help to add depth to decisions, in addition to making them more explicit and transparent, thereby enabling accountability for the choices that are made.

Given the pressure relating to trust in financial institutions and the replication skills of AI robots, the greatest impact could be achieved if AI Robots were not to focus on the best interest of clients, but on the profit of the company, thereby having a major effect on the company's reputation within society. It is therefore essential to analyze the manner in which moral and ethical decisions can be implemented in AI robots in order to ensure the best possible security against risks, while also allowing AI robots to help us improve our own moral and ethical decision-making.

Conclusion: Focus on framework

Analyzing ways in which to provide for meaningful moral and ethical decision-making requires a clear focus. In this project, the focus is on how to establish a "framework" with which to make decisions concerning moral and ethical questions. From a more future-oriented perspective, it also considers the impact of AI robots. The development of such a framework also requires serious attention to aspects of governance and leadership within the context of robot decision-making. More specifically, the focus of this Capstone project is as follows:

- **Main focus:** The moral and ethical decision framework
- **Secondary focus:** Individual leadership in the financial industry
- **Tertiary focus:** The governance of the moral and ethical behavior of AI robots

Based on the results of a literature study, the project proceeds from the framework developed by Norman Vincent Peale and Kenneth Blanchard, which contains a three-stage format.

- *Level 1: Is it legal?*
- *Level 2: Is it balanced?*
- *Level 3: How will it make me feel about myself?*

Based on these three simple questions, a Delphi study was conducted, in which experts were asked whether these questions are still fundamental, or if they should be adjusted to be suitable for the longer term, given the impact of AI robots.

Other innovative developments (e.g., blockchain) and other market segments are excluded (although the basic results might be useful to other markets).

As stated in the previous chapter, this project focuses on the following question: *What should be the moral and ethical decision framework in a digitalized financial industry?* The answer to this question is explored through qualitative research methodology. The most common qualitative research methods are participant observation, in-depth interviews, and focus groups. This research paper is based on in-depth interviews in the form of a Delphi study. The Delphi technique has five main characteristics (Bryant, 2011):

1. Focus on researching the future or topics about which little is known
2. Reliance on the opinion of experts
3. Utilization of remote group processes
4. Adoption of an iterative research process
5. Creation of a consensus of opinion

The value of the Delphi research outcome will depend largely on the quality of its input. The questionnaire that is used should be strongly based on a study of relevant literature.

The research comprises two steps. The first step consists of a literature study. Based on the outcomes of the literature study, the second step consists of applying the Delphi research methodology to input obtained from a variety of experts.

1. Literature study

The literature study involves the analysis of various publications that are relevant to the development of a moral and ethical framework. The study begins by examining the historical foundations of morality and ethics in relation to the framework. Three specific frameworks are addressed: the Utrecht Model, the dilemma model, and the three ethics check questions developed by Peale and Blanchard.

The study continues with an analysis of several different perspectives on Artificial Intelligence (AI):

- Risks and opportunities associated with AI
- AI and feeling
- AI in the financial industry
- The ethics of AI

The research results are presented in Chapter 4.

2. Delphi Method

The second research methodology is based on the Delphi Method (Helmer, 1951), which is explicitly intended to be used when investigating questions for which no simple answer is available. The roadmap for this methodology consists of gathering a variety of expert opinions in multiple rounds, in order to combine different views and consolidate them into a general expert opinion.

In general, Delphi study (named after the Delphi oracle) involves seeking the opinions of experts on a subject on which there is no consensus. The answers of all experts are fed back to each other

(anonymously) in several rounds, with the goal of reaching consensus on the questions asked. This encourages the experts to adjust their opinions based on the answers of the other experts. During this process, the differences between the answers are expected to decrease with each round, eventually converging on the "correct" answer. The process ends after several rounds, with the media determining the answer.

The method process requires at least three stages:

- *First stage*: A questionnaire consisting of open-ended questions is sent to several experts.
- *Second stage*: Depending on the answers, a brief questionnaire consisting of closed-ended or multiple-choice questions could be sent to the same experts.
- *Third stage*: All of the experts are confronted with their own opinions in comparison to those of the other experts (on an anonymous basis), and the outcomes are addressed in an open, one-to-one discussion.
- *Fourth stage*: All of the outcomes of the open-ended questions, closed-ended questions, and personal discussions are combined in the final analyses.

This process provides the opportunity to obtain broad perspectives from a variety of experts at the start, to confront experts with differing insights from the other experts, and to combine those insights together to produce a supported general outcome.

Experts

The Delphi study requires a varied combination of expertise, including in the areas of morality and ethics, AI, and the application of decision frameworks in ethics committees. The experts who participated in the Delphi study are listed below, categorized by their specific areas of expertise.

- *Experts on moral and ethical decision-making*
    Prof. dr. Lans Bovenberg (Erasmus Economics and Theology)
    Dr. Peter Gillies DBA

- *Expert on ethics committees*
    Jennifer Rietveld (participant in the Achmea Ethics Committee, also responsible for the development of the Achmea moral compass)

- *Expert on future developments in AI robots*
    Prof. Carlos Cordon

- *Expert on AI in practice (health sector)*
    Mike van Rijswijk

The Delphi research questions, answers, and interview results are presented in Chapter 5.

Introduction

This literature study begins by exploring the historical foundations of morality and ethics, as it relates to the framework. The second section examines three different moral and ethical decision frameworks: the Utrecht Model, the dilemma model, and the three ethics check questions.

In preparation for developing a moral and ethical decision framework that takes into account the impact of AI robots, the third section elaborates on several different perspectives on Artificial Intelligence, as input for the Delphi Research. These perspectives differ according to 1) the risks and opportunities of AI; 2) AI and feeling; 3) AI in the financial industry; and 4) the Ethics of AI.

1.   The historical foundations of morality and ethics

Morality is built into us as social beings.

In the 18th century, the publication of Adam Smith's *The Theory Of Moral Sentiments* (Smith, The Theory of Moral Sentiments, 1759) demonstrated how our moral ideas and actions are a product of our very nature as social creatures. Smith argues that this social psychology is a better guide to moral action than is reason (Smith, AdamSmith.org, n.d.).

As individuals, we have a natural tendency to look out for our own interests. This is merely prudence. As Smith further explains, however, as social creatures, we are also endowed with a natural sympathy—in contemporary parlance, empathy—towards others. As we grow from childhood to adulthood, each of us gradually learns what is and is not acceptable to other people. Morality stems from our social nature.

Even though we are self-interested, we must learn how to live alongside others without doing them harm. This is an essential minimum for the survival of society. If people go further and do positive good (i.e., beneficence), we welcome it. According to Smith, however, we cannot demand such actions in the same way that we demand justice.

The ideal must be that any impartial person, whether real or imaginary (i.e., Smith's "impartial spectator") would fully empathize with our emotions and actions. This requires self-command.

Smith argues that morality is not something new that we must develop. It is an inherent part of us, as social beings.

Logos, pathos, and ethos

Aristotle's Art of Rhetoric is an ancient Greek treatise on the art of persuasion, dating from the 4th century BCE. Rhetoric makes use of modes of persuasion (often referred to as "ethical strategies"), which are devices that classify a speaker's appeal to the audience. Aristotle divides these devices into three categories: *logos, pathos,* and *ethos*.

In *Moreel Leiderschap* [Moral Leadership] (Brenninkmeijer, 2019), Alex Brenninkmeijer concludes that moral leadership is not a given talent, but the result of knowledge, experience, and the courage to engage in conversations with others about values. In this process, making mistakes and learning from these mistakes are crucial to further development.

"In addition to being the ingredients of the oratorical art, as described by Aristotle, logos, pathos and ethos are inseparable parts of our human existence. In everything that I think, feel, and do, these three interdependent aspects play a crucial role in my expression. Responsible judgment requires contemplating how logos, pathos, and ethos can be related to each other in a particular matter and discovering how I can bring them together in a balanced relationship. This applies not only to me, but also to the other, who can arrive at a different balance." [translation from Dutch] (Brenninkmeijer, 2019, p. 66)

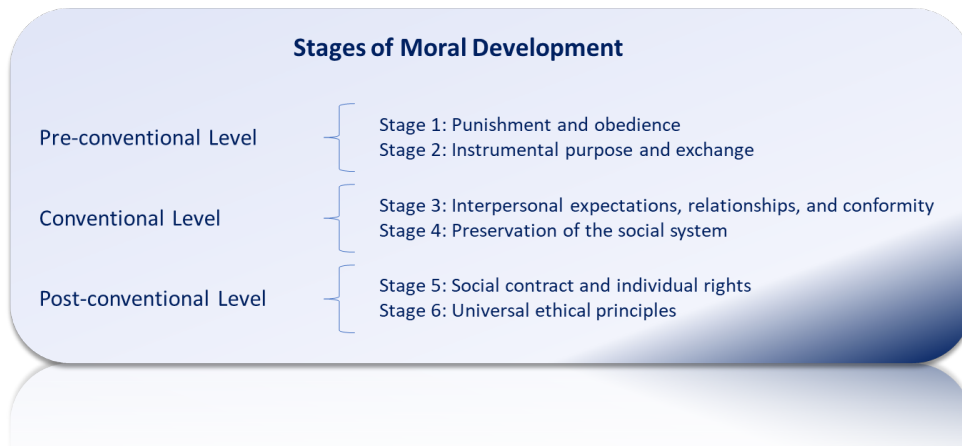Brenninkmeijer uses an equilateral triangle to illustrate the balance between the three factors.
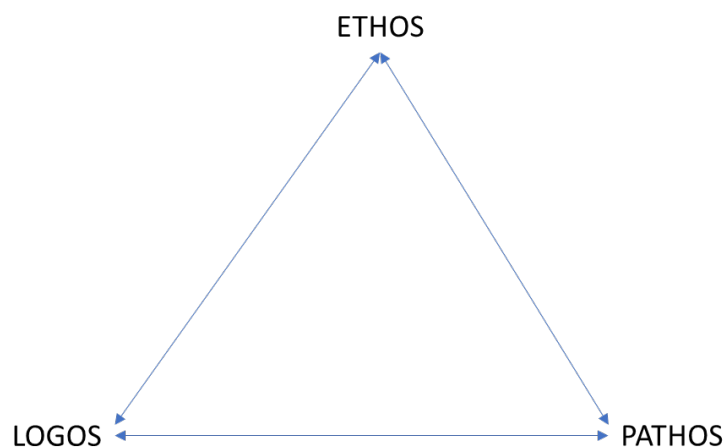
Self-awareness

The ability to conduct a proper dialogue on moral and ethical questions is not enough to answer them. Deeper knowledge of self-awareness is needed as well. For example, the Big Five personality model (extraversion, agreeableness, openness, conscientiousness, and neuroticism) could be helpful in this regard (E.C. Tupes, 1961).

As concluded by Kalshoven, Den Hartog, and De Hoogh (K. Kalshoven, 2011), and in line with available theory, conscientiousness, agreeableness, and emotional stability are positively related to ethical leader behavior. Requiring a standard Big-Five test of those who are responsible for moral and ethical decision-making could help to enhance self-awareness and develop moral and ethical leadership at a higher level.

The "moral stages" proposed by Kohlberg (see (Garz, 2009) are highly relevant to debates on morality and ethics . Kohlberg's six stages can be grouped into three levels of moral reasoning: pre-conventional, conventional, and post-conventional. According to Kohlberg, people can pass through these levels only in this order. Each new stage replaces the reasoning typical of the earlier stage. Not everyone attains all of the stages. Various aspects of this model of moral development have been criticized. For example, it has been argued that the model is culturally-biased and/or gender-biased and as being insufficiently empirical. It is nevertheless broadly acknowledged that Kohlberg's research is relevant, in that it enhances awareness and understanding concerning the various levels of moral reasoning.

Stages of Moral Development

Pre-conventional Level	Stage 1: Punishment and obedience Stage 2: Instrumental purpose and exchange
Conventional Level	Stage 3: Interpersonal expectations, relationships, and conformity Stage 4: Preservation of the social system
Post-conventional Level	Stage 5: Social contract and individual rights Stage 6: Universal ethical principles

The two systems proposed by Kahneman in Thinking, Fast and Slow (Kahneman, 2011) are also relevant with regard to moral reasoning. System 1 refers to the quick, intuitive way of thinking, and System 2 refers to the slow, thoughtful way of thinking. Moral illusions are likely to have the greatest impact in System 1. In an inaugural lecture (Bovenberg, 2019), Lans Bovenberg describes an even more specific way of thinking by combining "thinking, fast and slow" with "feeling, fast and slow" and "loving, fast and slow." The attempt to think, feel, and love slowly is essential to answering moral and ethical questions.

Most relevant insights to be used as input for the Delphi study:

- Morality is not something new that we must develop. It is an inherent part of us, as social beings.
- Moral leadership is not a given talent, but the result of knowledge, experience, and the courage to engage in conversations with others about values.
- Making mistakes and learning from these mistakes are crucial to further development.
- Responsible judgment requires contemplating how *logos*, *pathos*, and *ethos* can be related to each other in a particular matter and discovering how they can be brought together in a balanced relationship.
- Conscientiousness, agreeableness, and emotional stability are positively related to ethical leader behavior, and a high level of self-awareness is required in moral and ethical leadership.
- It is important to take into account various levels of moral reasoning and to focus on slow thinking, feeling, and loving, as our quick intuitive way of thinking is subject to the impact of moral illusions.

2. Frameworks of moral and ethical decision-making

The Utrecht Model, the dilemma model, and the three ethics check questions

The use of frameworks to answer moral and ethical questions in a systematic manner is common practice for ethics committees in the healthcare sector. Various methodologies are used, depending on availability of time, experience, and purpose. Examples of these methodologies include Socratic debate, the Dilemma method, and the Utrecht Model. Socratic debate is a form of cooperative argumentative dialogue between individuals, based on asking and answering questions to stimulate critical thinking and to draw out ideas and underlying presuppositions. The Dilemma method and the Utrecht model are more formal systematic approaches to answering a moral and ethical questions.

The Utrecht Model

The Utrecht model (Universiteit Utrecht, n.d.) is defined as the following process:

- First phase: Explication
  1. Which factual information is missing?
  2. What is the moral question?
  3. Which options for action are available?

- Second phase: Analysis
  4. Who is involved in the situation, and what is their perspective?
  5. Which considerations seem relevant?

- Third phase: Deliberation
  6. Which level of priority is assigned to each consideration?
  7. Which alternative seems preferable, based on the deliberation?

- Fourth phase: Implementation
  8. What seems to be the right thing to do?
  9. How will this be handled?

Dilemma method (van Dartel & Molewijk, 2014)

- Step 1. Formulate the dilemma
- Step 2. Describe value and norms
- Step 3. Search for alternative solutions
- Step 4. Make the deliberation

Although the Utrecht Model and the Dilemma Method are quite similar, the Utrecht Model provides a more detailed specification of the various steps to be taken and questions to be answered. In both methods, the crucial stage occurs in Step 4/the fourth phase, in which opinions clash and impactful debate starts.

Ethics check questions

Three ethics check questions are introduced in *The Power of Ethical Management* by Norman Vincent Peale and Kenneth Blanchard (Peale & Blanchard, 1988).

- Level 1: Is it legal?
  Will I be violating either civil law of company policy?

- Level 2: Is it balanced?
  Is it fair to all concerned in the short term as well as the long term? Does it promote win-win relationships?

- Level 3: How will it make me feel about myself?
  Will it make me proud?
  Would I feel good if my decision was published in the newspaper?
  Would I feel good if my family knew about it?

The most impressive features of this model are the simplicity of the questions and the interconnection of *logos*, *pathos*, and *ethos*, as described by Aristotle and debated in Brenninkmeijer's *Moreel Leiderschap* [Moral Leadership]: legality connected to *ethos*, balance connected to *logos*, and feeling connected to *pathos*. These three questions could also be helpful in the final stage of the Utrecht Model and Dilemma method, with the potential to help participants make their answers in this stage more explicit. Another advantage of the ethics check questions has to do with the possibility of using data to answer such questions as "Is it legal?" and "Is it balanced?" The capabilities of AI Robots are likely to be most helpful in the analysis of data.

How will it make me feel about myself?
In light of Smith's argument that morality is not something we must calculate, but an inherent part of us, as social beings, the "Level 3 'feeling question'" would seem to be most the crucial of the three ethics check questions to take into consideration.

Given that feelings are also the aspect that is most likely to be influenced by personal opinion, an additional re-check should be quite helpful. Maryam Kouchaki, an adjunct professor at the Kellogg Business School in Chicago, has developed more detailed questions for Level 3. These questions provide three concrete ways to test whether one is making the right decision (Ellemers, 2020).

- **The publicity test:** What if it were to be reported in the newspapers? What would your customers, employees, or family think about you making this decision?

- **The generalization test:** What would you think if everyone made the same choice as you, and what would the consequences be?

- **The mirror test:** When you make this choice, do you see in the mirror the person you want to be?

Most relevant insights to be used as input for the Delphi study:

- Use the experience of the healthcare sector concerning moral and ethical decision making as inspiration for the financial Industry. Debate the various methods, and establish an ethics committee for the financial institution, given that the number of moral and ethical questions will increase in response to future developments in AI.
- In the systematic approaches of the Utrecht Model and Dilemma method, the final phase is most crucial as personal opinions and weight are added to each argument and discussed with each other.
- The most impressive features of the ethics check questions are the simplicity of the questions and their interconnection to *logos*, *pathos*, and *ethos*.
- The three ethics check questions provide a more in-depth, systematic approach that could be used in the final stage of the debate. It also provides the possibility of using data to answer questions concerning legality and balance (and AI Robots themselves could prove useful).
- The feeling check question is most relevant to the moral and ethical framework, and the ethics tests provide the possibility of answering this question in greater depth.[3]

Given that the ethics check questions are relatively simple to use, are able to be supported by data (thus allowing the implementation of AI), and integrate all three aspects (*logos*, *pathos*, and *ethos*), it is used as the starting point for a decision framework to be challenged in the Delphi study.

---

[3] *The publication by Ellemers and study by Kouchaki were perfectly timed and integrated in Phase 2 of the Delphi study. The expert outcome was unanimous with regard to adding the three test questions as a guide for answering the feeling question.*

3. Artificial Intelligence

Risks and opportunities of AI; AI and feelings; AI in the financial industry; and Ethics in AI

3.1 Risks and opportunities of AI

Asimov's "Three Laws of Robotics" were presented in 1942 (Asimov, 1939-1950):

- *A robot may not injure a human being or, through inaction, allow a human being to come to harm.*
- *A robot must obey orders given it by human beings except where such orders would conflict with the First Law.*
- *A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.*

In their simplicity, these three rules provide an early projection of the possible impact of AI robots.

In "An experiment: The gods of cyberanthropology" Gillies provides an interesting analysis of Asimov's rules (Gillies, 2019). Gillies concludes that current developments in AI are already in violation of the Asimov Rules. For example, the Defense Advanced Research Projects Agency (DARPA) is using AI to harm humans (Asimov Rule #1), and our robotic servants have not been given the power to override orders that violate the stated rules (Asimov Rule #2). Stating the reasoning behind the article, Gillies writes, "…we have not – and probably will not – 'Purposefully Design' A.I. to the standards suggested by Asimov or business consultancies, philosophers, priests, or kings. […] We must endeavor to truly understand how we are developing the A.I. brain before we do something we regret – if we've not yet done it already" (Gillies, 2019, p. 6).

Gillies further states that "programming 'good' into A.I. assumes: 1) we know what it is and 2) we assume there will be no corresponding A.I. insight into evil. […] The chances that we will program 'good' into A.I. are just as high as programming in 'evil'. Without an exclusive transcendent ethical reference point A.I. will use its programmers as that reference point…" (Gillies, 2019, pp. 9, 10).

Gillies' final conclusion is that issues concerning anthropology should be framed according to "four main questions" (Gillies, 2019, p. 13):

1) *Meaning: What should be meaningful to A.I. and how do we make it aware of this meaning? – here we must understand what responsibility(-ies) A.I. should assume and should want to assume. The ox yearns to pull, the ass to be usefully burdened; what will A.I. "want"?*
2) *Sin: What superordinate value hierarchies should A.I. receive – or is it to build them itself? – what are the values – the "good" that A.I. perceives and what can be built in as undesirable? This echoes Asimov's work with robots, but it cannot be purely prescriptive and must, to some degree, be discovered by A.I. since we have missed our opportunity to "hard wire" its morality.*
3) *Pain: What is the equivalent of pain for A.I, and how should we build it in retroactively? – perhaps at a fundamental level. We have suggested that avoidance of pain can bring some sort of "meaning" for A.I. Pain and its alleviation may be a cornerstone of A.I. morality.*
4) *Death: What is the equivalent of death for A.I. and how should A.I. become aware of it? – pain and death may be our only protection against a rampaging A.I., but the moral landscape that accompanies it is treacherous and uncharted. (Gillies, 2019, p. 13-14)*

European Group on Ethics in Science and New Technologies

Since the publication of Asimov's rules, technology has developed rapidly and at an unprecedented level. Moral and ethical questions concerning AI robots have been the focus of considerable discussion, including within the European Commission (EC). In March 2018, the European Group on Ethics in Science and New Technologies (EGE) presented their *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems* (Technologies, 2018):

> *Advances in AI, robotics and so-called "autonomous" technologies have ushered in a range of increasingly urgent and complex moral questions. Current efforts to find answers to the ethical, societal and legal challenges that they pose and to orient them for the common good represent a patchwork of disparate initiatives. This underlines the need for a collective, wide-ranging and inclusive process of reflection and dialogue, a dialogue that focuses on the values around which we want to organise society and on the role that technologies should play in it.*

> *This statement calls for the launch of a process that would pave the way towards a common, internationally recognised ethical and legal framework for the design, production, use and governance of artificial intelligence, robotics, and 'autonomous' systems. (Ethics, 2018, p. 5.*

The EC has recognized the impact of AI, observing that many AI robots have been "released into the world unsupervised and may accomplish things which are not foreseen by their human designers or owners. [...] AI's inner workings can be extremely hard – if not impossible – to track, explain and critically evaluate." (Technologies, 2018, p. 7). They further observe: "a push for ever higher degrees of automation and 'autonomy'" in AI robots, noting that "investments of countries and large companies in this field are enormous and a leading position in AI research is among the prominent goals of superpowers in the world" and that "there is a development towards ever closer interaction between humans and machines" (p. 7). The EGE is concerned about these developments. In addition to the risks, however, the EGE has acknowledged the opportunities presented by AI robots.

> *Artificial intelligence, robotics and "autonomous" systems can bring prosperity, contribute to well-being and help to achieve European moral ideals and socio-economic goals if designed and deployed wisely. Ethical considerations and shared moral values can be used to shape the world of tomorrow and should be construed as stimulus and opportunities for innovation, and not impediments and barriers. and leads to more and more moral and ethical questions. (Technologies, 2018, p. 20)*

The development of a moral and ethical frameworks could not be more applicable to preparing the contemporary financial industry for tomorrow and beyond. The EC has identified this market as the third most important area of application of AI robots, noting that "A third important area of application is 'autonomous' software including bots. Trade, finance and stock markets are largely run by algorithms and software." (p. 11). If their actions are not supervised, AI robots thus have the potential to have a major impact in these contexts.

The ethical principles and pre-requisites developed by the EGE can provide a foundation for developing a moral and ethical framework for issues relating to the use of AI robots:

(a) Human dignity: *The principle of human dignity, understood as the recognition of the inherent human state of being worthy of respect, must not be violated by 'autonomous' technologies. [...]*

(b) Autonomy: *The principle of autonomy implies the freedom of the human being. [...]*

(c) Responsibility: *The principle of responsibility must be fundamental to AI research and application. [...]*

(d) Justice, equity, and solidarity: *AI should contribute to global justice and equal access to the benefits and advantages that AI, robotics and "autonomous' systems can bring. [...]*

(e) Democracy: *Key decisions on the regulation of AI development and application should be the result of democratic debate and public engagement. [...]*

(f) Rule of law and accountability: *Rule of law, access to justice and the right to redress and a fair trial provide the necessary framework for ensuring the observance of human rights standards and potential AI specific regulations. [...]*

(g) Security, safety, bodily and mental integrity: *Safety and security of 'autonomous' systems materialises in three forms: (1) external safety for their environment and users, (2) reliability and internal robustness, e.g. against hacking, and (3) emotional safety with respect to human-machine interaction. [...]*

(h) Data protection and privacy: *In an age of ubiquitous and massive collection of data through digital communication technologies, the right to protection of personal information and the right to respect for privacy are crucially challenged. [...]*

(i) Sustainability: *AI technology must be in line with the human responsibility to ensure the basic preconditions for life on our planet, continued prospering for mankind and preservation of a good environment for future generations. [...]* (Technologies, 2018, p. 16-19)

From implicit to explicit moral and ethical choices

In *Customers the day after tomorrow* (van Belleghem, 2017), Steven van Belleghem describes the future of AI and the impact (both positive and negative) in daily life. The benefits of AI for individuals, companies, and society can be realized through proactivity and debate with regard to the ethics of AI. Van Belleghem notes that, in the absence of such efforts, the future impact of AI is certain, given all of the risks that accompany it.

It is important to be aware of the technology filters that are currently used in our decision-making processes. One interesting finding of a study by Google, van Belleghem stated, is that the most searched question following BREXIT was "*What is the EU?*". Confronted with the outcome of a complex decision that was destined to have a major impact, individuals were most likely to focus their research on what the decision was actually about.

In the future, AI will force brands to make explicit ethical choices. Banks seeking to work with AI robots will have to make explicit statements concerning how they intend to balance their interests against those of their clients. Such explicit choices are not necessarily required for human-to-human interactions. The technological revolution will help prepare companies to make and communicate about these explicit choices. Transparency will also help companies to develop greater public trust.

One relevant question in light of the developments outlined above concerns who will lead this moral and ethical debate and ensure the required transparency. Although governments will obviously play a crucial role, large tech companies are international, and thus not subject to local restrictions. Van Belleghem advises the development of a collaborative model in which governments and companies work together in public-private efforts to promote higher levels of moral and ethical behavior.

3.2 AI and feeling

What are the human capabilities of Artificial Intelligence? In *The Second Machine Age* (Brynjolfsson & McAfee, 2016), Brynjolfsson and McAfee argue that AI is incapable of undoing one important aspect of human capability: : ideation. By definition, AI cannot conceive of new ideas or concepts. More precisely, it is incapable of coming up with good ideas.

According to Brynjolfsson and McAfee, although computers can easily be programmed to produce new ideas, all of their ideas are based on data. They could thus more accurately be referred to as "digital equivalents." As with the proverbial monkeys and typewriters, they could never reproduce a or create a single Shakespeare play.

> *"Ideation in its many forms is an area today where humans have a comparative advantage over machines. Scientists come up with new hypotheses. Journalists sniff out a good story. Chefs add a new dish to the menu. [...] Many of these activities are supported and accelerated by computers, but none are driven by them." (McAfee & Brynjolfsson, 2016, p. 191)*

It would seem logical for AI robots to be incapable of human "ideation." Another important question, however, concerns whether they might already have started to develop relationships, albeit based on data.

> *"Technology is not destiny, we shape our destiny"*
> *(McAfee & Brynjolfsson, 2016, p. 257)*

Robot Phi

Mike van Rijswijk is developing experiments with AI in the healthcare sector. One famous experiment involved Robot Phi, in cooperation with Stichting Philadelphia and The Innovation Playground.[4]
The most impactful occurrence in that study was the development of a relationship between a person with disabilities ("Lydia") and Robot Phi. Lydia's description of their relationship was human.
Van Rijswijk explains this example of friendship between humans and AI in a paper for Achmea: "*Innoveren vanuit het hart*" [Innovation from the heart] (Van Rijswijk & Teirlinck, 2020).

> *The next question is thus, "Can a robot actually be a friend, like a human, or even a loved one? Classic philosophy cannot help us here. Aristotle's classic definition of friendship presupposes a degree of reciprocity and communality that cannot be expected in a human-robot relationship. One interesting question, which was also posed by the scientist de Graaf, concerns whether this is the proper yardstick for measuring the impact and value of human-robot relationships. How can we measure the value of friendship? In this regard, society often seems to assume that, by definition, the value and impact of a human-robot relationship is less than that of a human-human relationship. From this perspective, the very possibility that relationships could arise between humans and robots could disrupt or even threaten relationships between humans. In relationships between humans and animals, however, no questions are asked regarding whether the perceived bond between humans and animals might disturb relationships between humans and humans.*

---

[4] *Philadelphia Zorg - Robot Phi stayed with Lydia for two weeks. See also: https://www.youtube.com/watch?v=qIKt1BgLi1Y*

How many people have described their four-legged friends or cats as permanent members of their families, whose death will be perceived as a very painful parting? Should we therefore consider the human-robot relationship as a separate dimension, which is not threatening, but is simply an addition to the spectrum of relationships that a person can currently have? [translation from Dutch] (Van Rijswijk & Teirlinck, 2020, p. 35)

Ensuring data privacy between human and robot is of the utmost importance, especially in the case of people with disabilities. In their paper, Van Rijswijk and Teirlinck describe several different perspectives on data privacy. One popular view is that data should be private, unless there are good reasons for infringing upon such privacy. For example, data breaches could help to make society safer (e.g., the interception of telephone conversations in order to prevent terrorist attacks). Another popular view on data sharing, the business value perspective (e.g., Apple, Facebook), is in direct contrast to the one described above. These two opposing views have been identified as the social-cultural and the market perspectives, respectively (Thévenot & Boltanski, 2006).

Critical and ethical questions have been asked ever since the initial developments toward the Internet of Things (IoT). In recent years, these criticisms have increasingly come from developers and designers themselves. Despite efforts by such entities as the European Commission, an ethical framework for the IoT has yet to be developed, as suggested in scientific studies by Weber in 2013 and by Baldini and colleagues in 2016. The purpose of an "ethical framework" or "ethical design" is that it allows the design to be built in such a way that individual users can determine their own ethical choices, instead of their being determined in advance by the company or developer who happens to write the code (Baldini, Botterman, Neisse, & Tallacchini, 2016). This guarantees the values of individuals, but not at the expense of technology as a whole. It is obviously easier said than done. The large number of stakeholders, the lack of technical knowledge among policymakers, and the current dominant position of such tech companies as Google or Amazon are making this a nearly impossible task. [translation from Dutch] (Van Rijswijk & Teirlinck, 2020, p. 40)

As concluded by Van Rijswijk and Teirlinck, even if companies were to feel a moral responsibility to restrict data sharing to the highest level of privacy, the designer would still bear responsibility for the end result of programming AI robots and their behavior. As demonstrated in cases like that of Robot Phi, the impact of AI robots could very closely approach the equivalent of a human touch in personal life.

3.3 AI in the financial industry

Like the EC, which has recognized the financial industry as an important market with regard to the potential impact of AI robots, the Dutch Association of Insurers has presented a moral and ethical framework on AI.

Given its focus on changing its morality, the financial industry has acknowledged that the profit margins of insurers have been under constant pressure due to low market interest rates in recent years. Self-learning computers could improve their ability to analyze the increasing volumes of data on customers and damages, while offering the possibility of increasing sales and improving the efficiency of business operations. New techniques could also allow insurers to develop more personalized products. Moreover, AI robots could enhance the ability to compensate for—or even prevent—damages and detect fraud. They could further enhance the efficiency of efficient product pricing based on big data. Despite the potential advantages, the use of big data and AI are also accompanied by risks and issues of trust. For example, what happens if the data entered into the systems are not correct? Would anyone be insurable if insurers were to know everything?

The misuse of AI techniques on the part of insurers could lead to discrimination and uninsurability. As the supervising body the Netherlands Authority for the Financial Markets (AFM) have warned:

> *"It is important for AI to be used in a manner that is responsible and that meets requirements regarding sound, controlled business operations, product development, and duty of care."* [translation from Dutch] ((AFM), 2019)

Insurers should be able to explain their techniques and their results to customers. The Dutch Association of Insurers has established an ethical framework (Verbond van Verzekeraars, 2020), which has been endorsed by 95% of all non-life and life insurers in the Netherlands. The framework is based on the principles of the European Group on Ethics in Science and New Technologies.

Although more focus on ethics in the insurance industry is welcome, and although the further development of the ethical framework will have a positive impact, it is interesting to note that the framework has yet to focus on any aspect of the processes through which decisions are made and justified. Its focus is more on outcomes than on challenging the ways in which moral and ethical questions are addressed.

Robo-advice

In light of developments in the ethical framework that has been endorsed by the insurers, developments in the field of "robo-advice" are particularly relevant (Maatman, 2019).

"When the New Pension Agreement is executed, the importance of the contribution agreement will increase. Participants will be presented with choices that could potentially have major consequences for their pensions. In order to make responsible choices, they will need good information and advice. The large-scale provision of such advice will require the use of robots and algorithms.

Given the massive number of pension participants, advisors will have to make use of robo-advice. With full robo-advice, a robo-advisor completely takes over the work of humans. All aspects—including the customer's personal financial position, risk appetite, objectives, knowledge, and experience—are

processed digitally, and the advice is produced and communicated automatically. Human input is limited to the development of the software that generates the robo-advice and the possibility of resolving malfunctions in IT systems. Robo-advice requires customer data. The robot's algorithm combines, compares, and analyzes various sources of customer information, and the quality of robo-advice increases with the volume of customer data. The AFM (Autoriteit Financiële Markten) emphasizes that robo-advisors must have access to complete and correct information. Otherwise, their advice is unreliable. Providers of robo-advice provider will have to organize their IT management in ways that guarantee the integrity and confidentiality of customer (and other) data and the availability of the IT system.

In addition to making factual decisions, robots sometimes implement them as well. Mass data serves as a fuel for algorithmic decision-making. At the same time, mass data are the result of the application of algorithms. The collection and processing of mass data generate new data, which in turn enhances performance and improves algorithms. Briefly stated, providers of robo-advice must ensure that their robots and algorithms serve honest behavior: not manipulative and not dominating behavior (which would restrict freedom). [translation from Dutch] (Maatman, 2019, p. 6)

Maatman concludes that algorithms are ultimately used to classify and influence people. Obtaining mass data provides power over others. Mass data can be used to do either right or wrong. Maatman further points to the existence of information asymmetry, in which users of robots and self-learning algorithms possess infinitely more knowledge and options for influence than do the customers who provide their personal information. Additional attention should therefore be devoted to outcomes for clients, as AI robots (and their designers) might not always be focused on their best interest.

With regard to the liabilities associated with AI algorithms, an article by Eric Tjong Tjin Tai (Tai, 2017) provides relevant insight into regulations for algorithmic trading in the United States.

In America, the Financial Industry Regulatory Authority (FINRA) imposes various requirements on companies that use algorithmic trading. Among other requirements, the FINRA requires that the design of the algorithm (including their monitoring and review) be performed under the responsibility of a registered individual. Furthermore, some guidelines concern "supervisory and control practices." These guidelines include a general risk assessment, documentation that allows others to understand the code without studying the actual code, mechanisms for disabling the algorithm quickly, extensive testing that also searches for unintended side effects, and the implementation of controls and warning signs that signal unintended effects. The Securities and Exchange Commission (SEC) prohibits exchange traders from providing their clients with "unfiltered" market access, meaning that they must monitor the orders given by their clients, including the prevention of mistakes. [translation from Dutch] (Tai, 2017, p. 9)

Tai also describes the requirement to include a "kill switch"—an ultimate possibility for shutting off an AI robot. The experts who were interviewed unanimously agreed on the importance of this aspect and the expectation that such options will be necessary in the implementation of AI.

3.4 Ethics in AI

The ethics of AI is an emerging area. The first fundamental ethical framework was developed by Asimov in the "Three Laws of Robotics". These early rules have been adapted into guidelines, as presented by the European Group on Ethics in Science and New Technologies and translated to the Dutch market by the Dutch Association of Insurers. They have thus developed over time, and they will be increasingly operationalized in the years to come.

In *Robot aan het stuur* [A Robot at the Wheel] (Eynikel, 2017), Eynikel defines five priorities for the development of an ethical standard and legal framework with a more operationalized focus.[5]

1. **Level of autonomy:** To what extent is AI able to make the decision fully automated?
2. **Adjusted liability:** The driver currently assumes all liability for self-driving cars. Should the company not be involved as well?
3. **Compensation agreements "without correction":** In the United States, a health-insurance system is active, in which the liability risks for new vaccines are limited for companies. This is intended to ensure that they will be able to continue in case of unforeseen effects should emerge (i.e., if risks were to increase to such a level that a single mistake could bankrupt a company). A similar system could be institutionalized for AI providers and their specific risks.
4. **Transparent algorithm, rules, and principles:** Robots should always explain their decisions. Although the drivers of cars need not become AI experts, they should know what their cars will do in dilemmas occurring in the most critical situations (e.g., saving the driver versus the person in front of the car).
5. **Sharing is caring:** A risk level of zero is not possible. For example, the risks of self-driving cars can be minimized, but not eliminated.

According to Eynikel, when a robot is at the wheel, it is because humans put it there. Humans are therefore the ones who decide where AI will and will not be active. As the impact of technology on our daily lives increases, our duty to exercise caution with regard to the values that are affected will increase as well.

> *"The technologies we are creating provide vastly more power to change the world,*
> *but with power comes great responsibly"*
> *(Brynjolfsson & McAfee, 2016, p. 256)*

In an ideal world, it would be possible to code ethical guidelines for AI-enabled systems that would ensure that such systems would automatically behave ethically. Such developments are far in the future, however, if they are possible at all. Within the current state of the market, it is more practical to teach developers how to develop ethical systems, as stated in the 2016 Gartner AI Ethics Report (Buytendijk, Sicular, Bretheno, & Hare, 2019).

As leaders in the field of data and analytics face increasing pressure to create revenue, reduce costs, and mitigate risk, however, discussions concerning data ethics beyond mandatory legal compliance often wane. These leaders typically have little or no formal training in data ethics. Even when they do, they may not have the support, time, or resources to make digital ethics a priority. The implementation and achievement of data ethics are not usually components of the metrics or measures of success used by leaders in the field of data and analytics. Instead, data ethics is usually treated as a "nice to have" or considered "after the fact" in response to negative press. This gap in

---

[5] *These priorities were used as input for the questions included in the Delphi study on AI.*

executive oversight is the primary conclusion presented in the 2018 Gartner AI Ethics Report (Clougherty Jones, Buytendijk, & Hare, 2018).

Some leading companies have chosen to organize discussions concerning ethics in AI more permanently by instituting AI Ethics Advisory Boards comprising a diverse set of internal and external stakeholders. Such institutionalization provides the opportunity to discuss moral and ethical questions and to make decisions more explicit. One of the largest companies in the world, Google, reported that it was able to operate the advisory board for the maximum period of "one week":

> *"It's become clear that in the current environment, ATEAC (Advanced Technology External Advisory Council) can't function as we wanted. So we're ending the council and going back to the drawing board. We'll continue to be responsible in our work on the important issues that AI raises, and will find different ways of getting outside opinions on these topics"*
> *(Kent, 2019).*

Most relevant insights to be used as input for the Delphi study:

- Advances in AI robots have ushered in a range of increasingly urgent and complex moral questions.
- Asimov's "Three Laws of Robotics" and the EGE Ethical principles and democratic prerequisites (including the re-developed version prepared by the Dutch Association of Insurers) highlight the possible impact of AI in our daily lives. Although these ethical frameworks are highly relevant, they do not focus on any aspect of how the decisions are made and justified.
- Although AI cannot create ideas in the way that humans do, it can closely approximate a human touch with regard to feeling. It will nevertheless use its programmers as a reference point.
- The technologies that we are creating provide more power, but with power comes greater responsibly.
- Although ethics in AI is becoming increasingly operationalized, the institutionalization of AI Ethics Advisory Boards can be difficult to sustain.

A Delphi study is explicitly useful for addressing questions for which no simple answer is available. In this study, five experts responded to the questionnaire, and all four target groups[6] were covered by the participating experts,[7] thereby generating a highly relevant overview of different opinions from different perspectives.

Phase 1: Questionnaire

The questionnaire is divided into three parts. The first part focuses on the development of a new framework. The second part concerns the usability of the framework by the board of directors or by the members of ethics committees, and the third part covers the governance of AI robots.

The analysis of the framework is based on the "ethics check questions" proposed by Norman Vincent Peale and Kenneth Blanchard in their 1988 publication.

The following case study provides a more detailed explanation of the three levels. It was used to assist the experts in answering the questions in the questionnaire.

Example case study:

In an insurance company a new product is launched for disability insurance. The premium for the individual policyholder is calculated by an AI robot, based on a complex algorithm and using deep-learning skills.

The ethical question in this case is focussed on premium differentiation calculated by the AI robot.

> **Level 1** (is it legal?): the AI robot is discriminating on "surnames". A person with a French surname receives a lower premium offer than a person with an African surname. Discrimination on surnames is not allowed by law, it's discrimination based on ethnicity[8], the robot should be reprogrammed to exclude this way of discrimination.

> **Level 2** (is it balanced?): the AI robot is calculating the premium, but is the premium balanced, is it solidary to all insured participants (young vs old, man vs woman, etc.)? But also, is the insured risk in balance with the policy premium?

> *Achmea uses a five stakeholder model; customers, employees, (business) partners, shareholders and society. Given these five different stakeholders, are all relevant aspects for the different stakeholders balanced in the premium differentiation? And what is than balanced? Should it be equal weighted? And if so, how can you calculate a "balanced" method for society?*

> **Level 3** (How will it make me feel about myself?): is for instance the price of the product in balance with the profit you earn? Or would you be proud if the profit numbers are published in the newspaper?

6 Experts on AI robot developments and on moral and ethical decision making, participants in ethics committees in the financial Industry, and an expert on the usability of AI in daily practice.
7 Prof. Carlos Cordon, Prof. dr. Lans Bovenberg, Dr. Peter Gillies, Jennifer Rietveld, and Mike van Rijswijk participated. At the start of the Delphi study, seven experts were asked to participate. Three additional experts were later asked to participate as well. Unfortunately, the COVID-19 pandemic impacted everyone's daily life, and not all of the experts were able to participate, despite their enthusiasm.
8 Based on Article 1 of the Dutch constitution

The questions asked were consciously designed to be broad and general, and the experts were asked to answer them as extensively as possible. The questions and the consolidated feedback are presented below. The consolidation process started with the analysis of all individual feedback and the development of the general statement, whenever possible. This consolidated statement was then assessed by each expert in the individual interviews, in order to ensure consensus on that specific feedback.

Main focus: Moral & Ethical Framework

Making a moral decision requires first an analysis of the specific moral issue, on a macro, meso and micro level, but also from a local vs universal perspective, backward and forward looking[9].

When the analysis is finished, the case is clear, we can enter the decision making phase. Which framework is useful when making the final moral and ethical decision?

1. The framework is referring to for example the "ethics check questions" of Norman Vincent Peale and Kenneth Blanchard.

   - *Level 1: Is it legal?*
   - *Level 2: Is it balanced?*
   - *Level 3: How will it make me feel about myself?*

   a. Are these questions still the fundamental questions to be asked when answering a moral question? If so or not, why?

   > Answer:
   > Based on the expert feedback, a small closed questionnaire (phase 2) is asked as on one of the most fundamental research questions the consensus was medium and additional questions were asked.
   >
   > **Consensus (medium):**
   > The questions are helpful and contain the fundamental aspects.
   >
   > **Differentiation (high):**
   > Different additional questions are asked by the experts:
   > - Should we use fair instead of balance?
   > - What about deliberately wrong-doing (can we exclude that in a group?)?
   > - Should we add the question: "How we can compensate people who suffer because of the decision made?"?
   > - What (kind of) weight is added to the different stakeholders in "balance" or "fairness"?

---

[9] As described in Professionele Ethiek [Professional Ethics] by Rob van Es (van Es, 2011).

b. Is the feel (level 3) question not already being used when answering the questions on level 1 & 2 (law isn't always black/white and balanced is already a subjective opinion). If so or not, why?

Answer:
Consensus (high):
Feeling is crucial to incorporate in answering moral questions, but is also subjective, even when you answer level 1 and 2 (except AI which will exclude "feeling" if the data is not biased with feeling aspects). Feeling is an additional type of knowing, intellectual and emotional knowledge are both important and should both be considered and weighted.

Differentiation:
None

c. Which question(s) has to be added when we implement moral and ethical decision making in an AI robot? Do we foresee an extra "risk/control" question added?

Answer:
Consensus (high):
Risk/control question; Is it biased? Is the goal achieved for all stakeholders? One expert stated: "We know that we have to develop a fence, but have no idea in which form … "

Due to a lack of trust in AI we have to check if the AI learning curve is developing in the correct manner based on datasets and algorithms. By doing so the AI will be a better decision maker based in the end on more data than a human can interpret, especially through it's fast self-learning skills. But the robot should explain its decisions, always.

Differentiation:
None

d. Can we "digitalize" level 1 by AI robots[10]? And if so, can we also implement AI for level 2 and 3 questions? Why of why not?

Answer:
Consensus (high):
Level 1 should be possible (although might be biased). Level 2 will be more difficult and level 3 not. This because a human touch is always needed on level 3 and there is not enough data yet for level 2. The AI robot will be more and more helpful on assessments of level 2 & 3 as more data become available.

Differentiation:
None

---

[10] *Machine learning is already being used for court decisions in the United States, with research results claiming accuracy of 70%–79% (Faggella, 2020).*

2. Starting point of the analysis is the "ethics check questions" framework. But do we have to use questions to secure the correct decision making of moral questions? If not, what different approach(es) do you foresee?

    Answer:
    Consensus (high):
    The questions are helpful (maybe incorporate scenario discussions), but especially on level 2 & 3 it's most important to ask the right questions, which AI can't do. Although, with a standardized extensive set of questions AI is capable to make an assessment on level 2 & 3 answers. But the question will always be how to "weight" the values on level 2 as unbiased input for the AI to make the correct assessment. And can data be unbiased? And from which perspective is it unbiased?

    Differentiation:
    None

Secondary focus: Leadership

Suppose we were to adopt the framework of the "ethics check questions" in practice. What should the board of directors or the members of the ethics committee who will be making the decisions take into account?

1. When we ask the level 2 questions (is it balanced), do we have to implement a multi-level stakeholder model so all different views / stakeholders (e.g. focus on customers, employees, (business) partners, shareholders, society, etc.) are incorporated? Which stakeholders are relevant?

   Answer:
   Consensus (high):
   Yes, a multi-level stakeholder model has to be implemented on level 2. Take also into account the hierarchy and weight added to each stakeholder (the more data available the better AI is capable of taking into account different stakeholders).

   Differentiation:
   None

2. Level 3 is based on "feeling". Our feeling could be largely influenced by moral illusions[11]. How are we able to exclude moral illusions in our decision making?

   With regards to "feeling" we have to be aware of the "Thinking Fast and Slow", two systems of Kahneman. System 1 is our quick intuitive way of thinking and system 2 the slow thoughtful way of thinking[12]. System 1 is where moral illusions could have the most impact.

   Answer:
   Consensus (medium):
   Feeling is crucial in moral and ethical decision making. To exclude moral illusion the freedom to speak and appreciation of your personal input in the dialogue is required, but experts are convinced that it is still very difficult to exclude.

   A coherent world view will be helpful. Feelings should also never be considered in isolation. Have various people with different value systems complement each other. Have people explain their emotional considerations to others, and have these considerations critically reviewed by others. Give everybody time to reflect on one's personal feelings.

   In addition to thinking fast and slow, we also have feeling fast and slow, it is crucial to combine both systems in moral and ethical decision making.

   One of the experts described a coherent worldview as – something, you might say, that is 'written on your heart'. We can recognize the worldview we have or test by:
   i.   Listening to your intuition – what is the right thing to do?
   ii.   Balance your intuition with science.
   iii.   Bring rigor to your decisions through understanding decision logic (Coherent Philosophy).
   iv.   Look back at what has happened before: history.
   v.   Compare revealed literature.

   Differentiation:
   None

---

[11] As described in the book Morele illusies [Moral Illusions], by Stijn Bruers (Bruers, 2017). For example, consider the "tram ethical dilemma."
[12] In his inaugural lecture, Bovenberg (Bovenberg, 2019) propose an even more specific way of thinking: "Thinking, Fast and Slow," "Feeling, Fast and Slow," and "Loving, Fast and Slow."

3. Do you think that when you enable input from different cultural perspectives, diversity and inclusiveness, this will help moral and ethical decision making on a higher level[13]? Do you see differences between including different cultural perspectives vs diversity vs inclusiveness? And if you think it's helpful, in which way could we secure culture, diversity and inclusiveness on the highest level?

> Answer:
> **Consensus (high):**
> Yes, it is absolute crucial to take into account culture, diversity and inclusiveness in decision making. Also a coherent worldview is required and a special focus is needed on the weakest values.
>
> Culture, diversity or inclusiveness might refer to very specific biases that some societies have had for some time. AI is not bound by that, but it might develop biases against some parts of society.
>
> **Differentiation:**
> None

4. In which way could leadership justify their decision making? E.g. reporting for instance in accordance with the three ethics questions?

> Answer:
> **Consensus (medium):**
> Document the process of decision making (documentation on each level is maybe too much) when humans are making their decision. In the AI world it is not about documentation, it is therefore crucial to learn on an equal level as the AI is doing.
>
> **Differentiation:**
> None

---

[13] *"McKinsey has been examining diversity in the workplace for several years. Our latest report, Diversity Matters, examined proprietary data sets for 366 public companies across a range of industries in Canada, Latin America, the United Kingdom, and the United States" (Hunt, Layton, & Prince, 2015, p. 2). See also "Is Moral Diversity a Good Argument for Moral Relativism?" (Melé & Sánchez-Runde, 2013).*

Tertiary focus: AI robots

Suppose that we have defined a framework and that our ethics committee has reached a decision on the moral and ethical question. How will we secure the input to and outcome from the AI robot?

1. What do you think is the impact of the AI robot in the financial industry?

> Answer:
> **Consensus (high):**
> The impact is huge, it will change the financial services completely from n=?? to n=1. AI is already used in several services, decisions are already based on lots of data (instead of feeling) and we even need more to analyse reactions.
>
> **Differentiation (small):**
> Trust is still the relevant factor for the financial industry, which cannot exclude a human touch. Be aware of the Tech companies that will enter the financial industry market (Google, Facebook, etc.).

2. How does robotization in the financial industry impact moral and ethical decision making?

> Answer:
> **Consensus (high):**
> The main challenge is to make sure that the decisions are not biased, that they follow ethical principles and are limited by boundaries, especially as the impact of an AI robot will exponentially get bigger.
>
> **Differentiation (small):**
> Focus on the individual will create more focus on excluding "bad risks". AI creates more insights and will be helpful in better moral and ethical decision making.

3. When the correct moral and ethical decisions are made and implemented in the AI robot, how can you secure the outcome if the robot is "self-thinking / non-scripted"[14]?

   a. In the implementation phase?
   b. And when the AI robot is activated and self-learning by external influences?
   c. How can you be aware of side-effects of the robot, unintended and unanalysed effects by dealing with an AI robot?

> Answer:
> **Consensus (high):**
> Develop a "red button" to stop the AI. As the AI is learning, we also have to keep learning how to secure the outcome, even develop specific AI to secure the outcome (the 2nd AI will become an expert in securing / controlling the AI outcome).
>
> The robot must always explain its decisions in an understandable manner for humans.
> We will need undiscovered fences.
>
> **Differentiation:**
> None

---

[14] *Non-scripted means that the input and output are not programmed upfront, but that the robot has deep-learning skills and will develop on its own.*

4. In which way should the standard PDCA cycle be adapted to the new situation, given the uncertainty with regard to the stability of the "plan," "do," and "check" steps, due to the fact that the robot is learning on its own.

> Answer:
> **Consensus (medium):**
> Current PDCA cycles are not capable of managing AI robots.
>
> **Differentiation:**
> Add the question "Does it still feel ok?". Develop new test systems. Develop new decision points. If the robot is learning, the checks and balances must develop equally.

5. Given the fact that the outcome of unscripted AI robots is unsure (as it is self-learning) we should take into account that errors will occur for sure. What is required on leadership if we have to deal with errors from the AI Robot?

> Answer:
> **Consensus (medium):**
> Errors will occur, you can't exclude them. Leadership should be capable to understand AI and learn equally on the capabilities of AI and be open and transparent on errors.
>
> **Differentiation (small):**
> Combine old (non AI knowledge) and new thinkers (AI believers) in teams and leadership. Different AI models are already used to check on errors.

6. Can level 3 (feeling) be implemented by a robot? Or is this always the human touch?

> Answer:
> **Consensus (high):**
> In the end a human touch is always needed.
>
> **Differentiation (small):**
> AI will already be helpful in assessments of decisions on level 3 nowadays, or it will be more helpful in the future: but a complete take over is not expected.

Phase 2: Multiple-choice questionnaire

In Phase 1, an extensive questionnaire was answered. The level of consensus was relatively high on most of the answers. The level of consensus on the most crucial question (e.g., *Are the "ethics check questions" which are used as a starting point for the analyses still the fundamental questions?*) was relatively low, however, and mixed feedback was received, in addition to the introduction of various important new aspects. For this reason, a "closed-ended questionnaire" was added in Phase 2, before starting the open discussion in Phase 3.

Question 1

Should we start the ethics questions with level 1 "is it legal?", or is it even better to start in the opposite order and start with: "How will it make me feel about myself?"

So re-arrange the questions in the following order:

- Level 1: How will it make me feel about myself?
- Level 2: Is it balanced?
- Level 3: Is it legal?

In answering the level 1 & 2 questions feeling is incorporated. It makes sense as some experts responded, to start therefore with level 3. Also, your feeling is crucial in answering moral and ethical questions and should be the starting point (the balance and legal questions function more as a check if the feeling answer is acceptable).

Answer 1

- Yes, start with the question: "How will it make me feel about myself"
- No, start with the question: "Is it legal"

Question 2

Should we choose to use the word "balanced" or is the word "fair" better? Which results in the following questions:

- Level 1: How will it make me feel about myself?
- Level 2: Is it fair?
- Level 3: Is it legal?

Balanced is more mathematical and fairness is harder and will therefore include a wider scope. But fair could be more subjective than balanced.

Longman Dictionary:

- *Balanced "giving equal attention to all sides or opinions"*
- *Fair "a fair situation, system, way of treating people, or judgement seems reasonable, acceptable and right"*

Answer 2

- Yes, change the word "balanced" into "fair"
- No, keep the word "balance" and do not use "fair"

Question 3

Should we add an extra question: "How we can compensate people who suffer because of the decision made?" Or can/should this be incorporate as one of the (multi-)stakeholder in "fair" or "balanced" on level 2

Answer 3

- Yes, adding an extra questions is preferable
- No, you can incorporate this focus in level 2 as one of the stakeholders

Question 4

Maryam Kouchaki, adjunct-professor at the Kellogg Business School in Chicage, provides three concrete ways to test whether you are making the right decision in her recent publication.

- ***The publicity test:*** *what if it gets in the newspaper? What would your customers, employees, or family think about you making this decision?*

- ***The generalization test:*** *what would you think if everyone made the same choice as you, and what would the consequences be?*

- ***The mirror test:*** *when you make this choice, do you see in the mirror the person you want to be?*

These three questions could be a better substitute for the single level 3 question "How will it make me feel about myself?" as the single questions is only focused on the person itself and these three questions have a broader focus.

Answer 4

- Yes, replacing the single question into three questions will be helpful in better moral decision making
- No, replacing is not necessary

Respondents results (Phase 2)
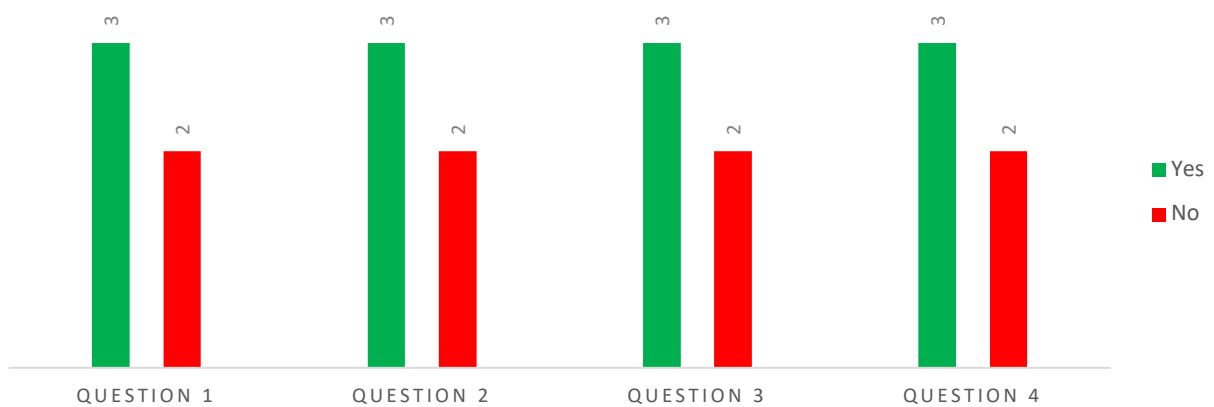
Phase 3: Expert interviews

All of the experts were confronted with their opinions in comparison to the other opinions, and the outcomes were finalized in an open discussion (one-to-one).

The answers presented in *Phase 1* are the final answers, as concluded from the initial personal feedback provided by the experts and the challenge to their answers in comparison to input from the other experts.

The interview started with the analyses of the multiple-choice answers provided in *Phase 2*. For Questions 1 & 2, the outcomes from the experts were not unanimous, and they differed when discussed in the interviews. All of experts were convinced that starting with feeling would be the best way to approach the process of answering moral and ethical questions, and they agreed that feeling is thus likely to influence the answer to Question 2. Starting with feeling would eliminate the necessity of adopting a broader focus on "fairness," which is already incorporated into the fundamental feeling check questions. Although experts agreed that *fairness* has more meaning than *balanced*, when something is equal (balanced), it is not necessarily fair. Particularly when assessments are developed by AI, it would be advisable to keep *balanced*, as this word is more capable of assessment according to data than *fairness*, which cannot be assessed in this manner. In addition, starting with the feeling questions makes it possible to challenge current legislation in terms of whether it is still correct or whether it should be adjusted. In the opposite order, current legislation is the starting point and is not challenged (this is especially relevant, as AI developments will obviously have an impact on current legislation).

The experts were also unanimous on Questions 3 & 4, following deliberation. With regard to Question 3, after deliberating on their thoughts, is the experts concluded that additional questions should not be seen as an additional "level," but as specific questions to be added when analyzing the perspectives of different stakeholders. It is especially important to take into account and address those who will suffer from decision. The question suggested in Phase 2 focuses on taking into account the compensation to this specific group. It is important to note that the focus should not necessarily be on actual compensation itself, but on the awareness of the views held by those who will suffer from the decision, with the goal of gaining a better understanding of their true feelings.

Question 4 is well-suited as an extension of the single, focused question on deeper thoughts divided in three different perspectives. It would be most helpful as a form of guidance for answering the feeling question. The problem with the first test (publicity) is that the answer depends on how another party views the matter. It should therefore be used only as a test. It should not be leading in terms of personal judgement, which should be the focal point. The three tests could be helpful in testing this. We should be acutely aware that the personal opinions of individuals are usually strongly influenced by others. Moral and ethical questions are difficult, and influence by others will not always help individuals to make up their own minds.

"The AI system will make errors when learning according to the behavior on five airplanes. For thousands of airplanes, however, AI would learn much more than a human could imagine. The most important aspects is therefore that the "n" should be large enough. In the end, however, when mistakes happens, humans will more forgiving of those made by humans than they will be of those made by AI."[15]

---

[15] *Interview with Carlos Cordon on July 1, 2020*

The overall conclusion from the literature study and the Delphi study is clear. The "ethics check questions" are helpful as a starting point for making moral and ethical choices, but they need to be adjusted to fit future developments.

Based on the literature study and the expert opinions obtained in the Delphi study, it is crucial to take the following aspects into account:

1. The order of the questions should be reversed. The starting point should be Level 3 ("*How will it make me feel about myself?*") instead of Level 1 ("*Is it legal?*").

   The reasons for this change are as follows:
   - Level 3 is fundamental in moral and ethical decision-making.
   - Levels 1 & 2 are biased by feeling (Level 3). In analyzing whether a decision is legal or balanced, such bias should be either eliminated or explicitly taken into account.
   - Level 1 could be performed by an AI robot, thereby eliminating bias if the data is unbiased. If biased datasets or biased algorithms are used, however, AI will implement biased opinions. The question that remains concerns whether data can be unbiased and, if so, from which perspective would it be unbiased.
   - Conclusions at Levels 1 & 2 should be excluded if Level 3 is not incorporated. Reversing the order of the questions would ensure that feeling is always included. In the opinions of the experts, this aspect would never be taken into account by an AI robot.
   - Starting with the feeling questions makes it possible to assess whether current legislation is still correct, or whether should be adjusted. In the opposite order, current legislation would be the starting point, and it could not be challenged at all.

2. Based on the reversed order, Level 2 and Level 1 could performed by AI as an assessment in Stage 1. The experts were convinced that AI will ultimately be able to take over decisions concerning "Is it legal?" and "Is it balanced?", on the condition that all necessary data are available (and, if possible, unbiased).

3. Let answering the feeling question be guided by the three feeling checks: the *publicity, generalization, and mirror tests*. These questions are well-suited as an extension of the single, focused question on deeper thoughts divided across the three different perspectives. Personal judgements should nevertheless always be the focal point.

4. Be aware of the weakest stakeholders or those who will suffer from the decision.
   - Develop a multi-stakeholder model that takes into account the hierarchy[16] and weight assigned to each stakeholder.
   - Include an additional re-check question to ensure that those who will suffer from the decision are equally incorporated (focusing not necessarily on compensation, but on awareness of their views).

5. When finalizing the three questions, a re-check should be incorporated if the outcome is still aligned with the answer emerging at the start.

---

[16] *The aspect of power is mentioned in the Delphi-study from the perspective of hierarchy. There is weight added to each stakeholders, the more power one has could lead to a lower weight, to secure those who suffer from the decision are on an equal basis incorporated.*

The result is as follows:

Starting point—feeling is crucial: It is crucial to incorporate feeling into the process of answering moral questions, but it is also subjective. Feeling is an additional type of knowing. Intellectual and emotional knowledge are both important factors in moral and ethical decision-making, and they should both be considered and weighted. The *three feeling tests* offer a broad scope for analyzing moral decisions from a variety of perspectives. AI cannot check feeling and ask the correct questions. It is capable of assessing Level 2 (*"Is it balanced?"*) and Level 3 (*"is it legal?"*) only if all relevant data are available and unbiased. To minimize biased opinions and results, we should include a combination of slow/fast thinking, slow/fast feeling, and diversity and inclusivity in our decision-making.

Second level—stakeholders: It is important to incorporate a multi-level stakeholder perspective (e.g., customers, employees, business (and other) partners, shareholders, and society). We must assign a weight to each stakeholder, however, thereby taking into account differences in hierarchy. From the perspective of AI, assessment at Level 2 could be implemented in the short term if all relevant data are available. The data used should be as unbiased as possible, and they should lead to a fair outcome. If the availability of more data could change the outcome and provide a better moral and ethical decision, this should be taken into account. In the long term, AI should be capable of completing second-level decision-making processes.

Third level—legality: Once feeling and balance have been assessed, the next test concerns whether the decision is legally acceptable. Level 3 could be completely performed by AI within a relatively short term. Early examples from the United States have indicated that accuracy of 70%–79% is already feasible, based on the current dataset (Faggella, 2020).

The final test concerns whether the outcome at Level 3 is still in accordance with the decision emerging at the start.

AI will be helpful in moral and ethical decision-making. In Stage 1, it is capable of making assessments at Levels 2 & 3. In Stage 2, it could completely take over decision-making at these levels. The automated system will ultimately do an even better job, based on more data. Nevertheless, humans should always analyze the final outcome.

This Capstone project started with an example involving Mercedes and its self-driving cars. It concludes with an updated framework for answering the questions that will emerge tomorrow and beyond. The timelines and relevance of such a framework can be illustrated in a recent decision made by IBM.

> *IBM will no longer offer, develop, or research facial recognition technology* (Peters J. , 2020)
>
> *Facial recognition software has improved greatly over the last decade thanks to advances in artificial intelligence. At the same time, the technology — because it is often provided by private companies with little regulation or federal oversight — has been shown to suffer from bias along lines of age, race, and ethnicity, which can make the tools unreliable for law enforcement and security and ripe for potential civil rights abuses.*
>
> *IBM Statement: "IBM no longer offers general purpose IBM facial recognition or analysis software. IBM firmly opposes and will not condone uses of any technology, including facial recognition technology offered by other vendors, for mass surveillance, racial profiling, violations of basic human rights and freedoms, or any purpose which is not consistent with our values and Principles of Trust and Transparency. We believe now is the time to begin a national dialogue on whether and how facial recognition technology should be employed by domestic law enforcement agencies."*
>
> *"Artificial Intelligence is a powerful tool that can help law enforcement keep citizens safe. But vendors and users of AI systems have a shared responsibility to ensure that AI is tested for bias, particularity when used in law enforcement, and that such bias testing is audited and reported." (Krishna, 2020)*

IBM made the conclusion to stop the development of facial recognition software. In the context of this project, relevant questions concern how they arrived at a decision on this moral and ethical question. Was it based on the fundamental feeling checks? Was it unbalanced? Was it legally not allowed?

Based on the literature study and the Delphi study, the conclusion is as follows:

Artificial intelligence has a major impact. Its connectivity to other AI and its replication skills are unprecedented. It will develop further over time. Nevertheless, all of the experts reached the same conclusion: AI is incapable of assessing the question of "*How will it make me feel about myself?*". The literature study did not yield a different conclusion. More importantly, the experts agreed that feeling is a crucial element in decision-making concerning moral and ethical questions. It should be the starting and concluding point for any discussion on moral and ethical decision-making. For this reason, the questions should be turned "upside down," such that they start with the fundamental question based on feeling. This should nevertheless not be done solely from the perspective of the individual person. Instead, it should be guided by the three ethics-related feeling tests: 1) the publicity test, 2) the generalization test, and 3) the mirror test. It should also not be used as a lead for stand-alone results based on feeling, but should always be integrated with the answers to questions concerning balance and legality. It is the interconnectedness of the three levels that provides the best results when answering moral and ethical questions.

As concluded by Kahneman (Kahneman, 2011), there is a difference between fast and slow thinking. Bovenberg (Bovenberg, 2019) made this even more specific, noting a difference between slow/fast

thinking and slow/fast feeling. Gillies (Gillies, 2019) concludes that both (and even all four) of these aspects should be taken into account when completing the three ethics-related feeling tests.

The three ethics questions proposed by Peale and Blanchard (1988) could be seen as "building blocks" that could be stacked on top of each other, but also that this should be changed. Given the influence that our feelings have on the answers to the questions of "*Is it legal?*" and "*Is it balanced?*", feeling should be seen as the fundamental starting point, and with the tests for balance and legality being used to determine whether the decision based on feeling is balanced for all stakeholders and legally allowed. Additional attention is needed for those who will suffer from the decision, in order to ensure that their interests and feelings are sufficiently taken into account, not necessarily in terms of compensation, but with regard to awareness of their views.

In the near future, AI will be able to make assessments at all levels, completely taking over the process starting with the question "*Is it legal?*", provided that all of the required data are available and unbiased (meaning that the data will not automatically lead to a biased outcome). It is important to note that all data are biased from some perspective. We must nevertheless make these biases known in order to ensure the incorporation of correct data.

As the conclusion of this Capstone project, we present a new moral and ethical framework. A high-level version is presented below, with a version containing greater detail on the next page.

The moral and ethical decision framework

Level 1:
How will it make me feel about myself?
Guided by: publicity, generalization, and mirror test.

Recheck:
Alignment level 1

The moral and ethical question

Level 3:
Is it legal?

Level 2:
Is it balanced (for each stakeholder)?
Attention should also be paid to those who will suffer from the decision.

Tomorrow and beyond

The development of a moral and ethical decision framework including the impact of AI is quite interesting in conceptual terms, but it should be tested and evaluated in daily practice (e.g., in ethics committees in the financial industry). A follow-up to this research involving results from concrete applications on the usability of the framework would be highly interesting and more than welcome.

The moral and ethical decision framework

Level 1: The fundamental check

How will it make me feel about myself? Is guided by the three feelings check tests:

1. **The publicity test:** What if it were to be reported in the newspapers? What would your customers, employees, or family think about you making this decision?

2. **The generalization test:** what would you think if everyone made the same choice as you, and what would the consequences be?

3. **The mirror test:** when you make this choice, do you see in the mirror the person you want to be?

To minimize biased opinions, be sure to include thinking & feeling fast and slow along with diversity & inclusivity.

Level 1 cannot be performed by AI. In the end, a human touch is always needed.

Level 2: Balanced check

Is it balanced (for each stakeholder)?

Take the hierarchy and weight added to each stakeholder into account as well.

Examples: Customers, employees, (business) partners, shareholders, and society

This level (Stage 2) could be performed by AI, if data are sufficient and unbiased. Stage 1 uses AI as an assessment for Level 2.

Given the focus on outcomes/developments, additional attention should be paid to those who will suffer from the decision.

The moral and ethical question

Recheck: Alignment

Is the final outcome still aligned with the fundamental Level 1?
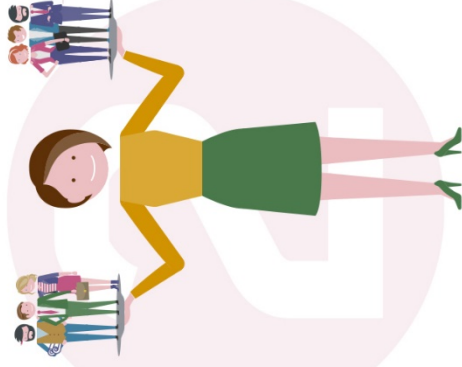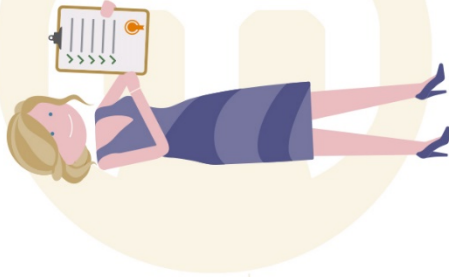
Level 3: Legal check

Is it legal?

The final check concerns whether the answer is allowed by law.

This level (Stage 2) could be performed by AI, if data are sufficient and unbiased. Stage 1 uses AI as an assessment for Level 2.

As a young boy, I dreamed of sailing on the ocean. I would later become a Maritime Officer and sail on the cruise ships of the Holland America Line. At a relatively young age, I became father and, standing on the bridge, I decided to take a new job, on land. In 2000, I started working in the financial industry, with a focus on pensions. Since 2003, I have had the privilege of working at Achmea, a company with a long-term focus based on its cooperative roots.

Working in the financial industry, especially in the front line dealing with clients with large assets, has provided me with insight into the culture and mindset in which self-interest and a short-term focus, bonus-driven culture continue to be relevant in daily operations.

As a Christian steward, I recognize God as the source of everything, and I know that we will ultimately be held accountable for the ways in which we use the gifts with which He has entrusted us. Although the financial industry plays a crucial in our daily lives, we must make a difference with regard to moral and ethical leadership, resolving not to be pushed by the outside world, but by our own intrinsic motivation. My personal focus in recent years has been directed more toward moral and ethical leadership: how can I make the difference? This focus has also been challenged in the E-MBA program at TIAS, and it has provided me with confrontational new insight.

Everyone is convinced that moral and ethical leadership is crucial, but no one could explain exactly how they make decisions on moral and ethical questions.
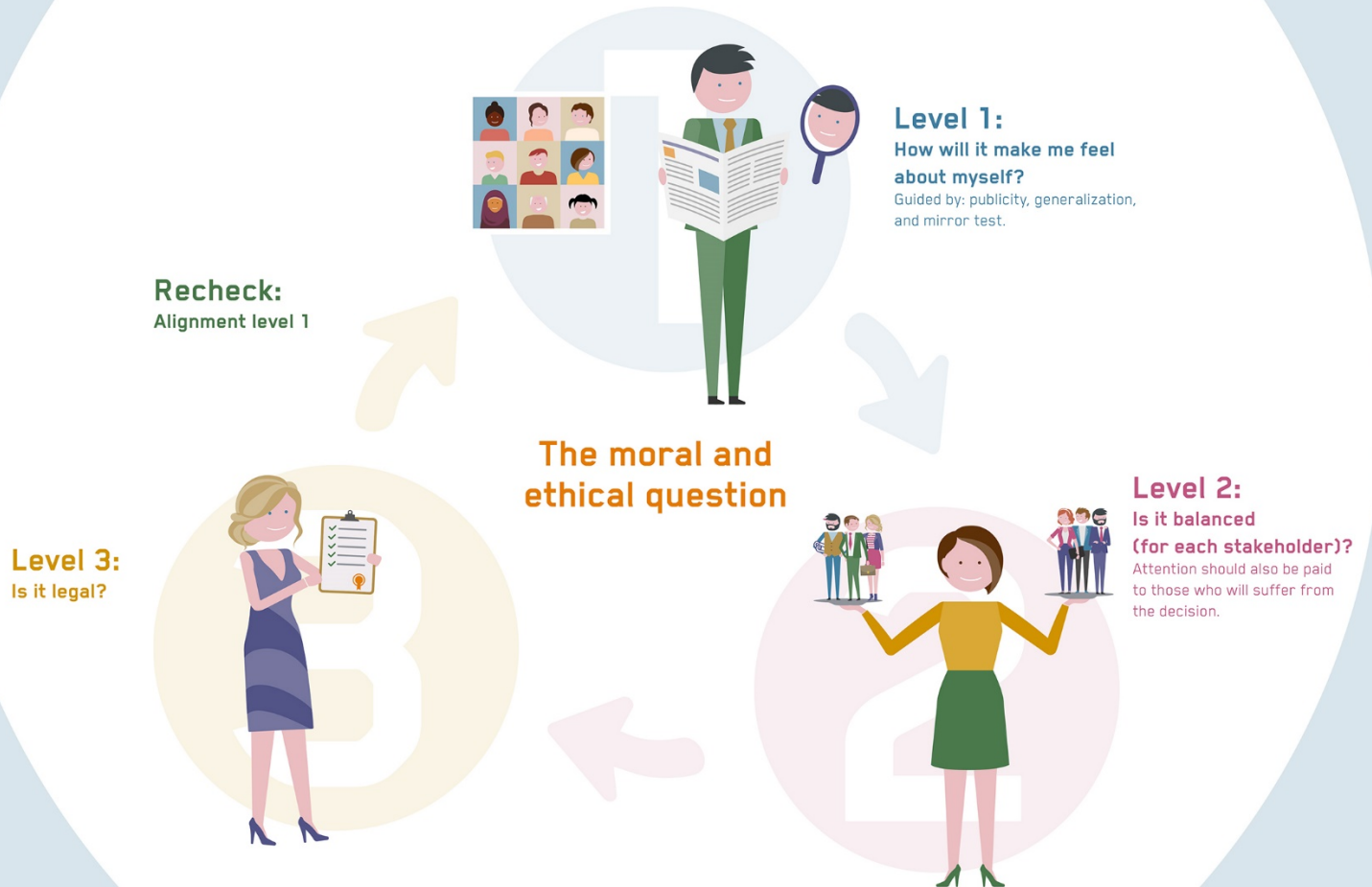
Moral and ethical discussions take place in special ethics committees, in boards of directors, in supervisory boards, and in other settings. Nevertheless, a structured manner of making such decisions has yet to be developed. Such a framework is also not presented or used as a subject for discussion in the curriculums of E-MBA programs and schools of economics. Why? One could argue that moral and ethical leadership and decision-making cannot be done in a prescriptive format. Although this is fundamentally true, it can never be captured in a checklist. At the same time, however, a conceptual framework could be quite helpful in challenging and explaining the decisions that we make as individuals and as groups.

In the not-too-distant future, I envision a world in which robots help us in our daily lives. I am following these developments with great interest. I have become aware that, despite their major impact, the development of the digitalized industry and AI has remained unrecognized, especially in combination with moral and ethical leadership. The implementation of AI in the financial industry could easily be at odds with moral and ethical decision-making. For leaders in the financial industry, it is important to acknowledge this and act on it in a proactive manner.

With this Capstone project, I have attempted to provide some preliminary answers for a moral and ethical framework. Working on it has helped to improve my understanding of moral and ethical decision-making, and it has provided me with new insights. It has also inspired me, as well as many others with whom I have had the privilege of discussing this subject.

I would like to thank all of the people who have been willing to listen to the concept, to debate it, and to make it even better, especially Eric Dooms. I am also grateful to everyone who wrote the books that I have read (some are mentioned in the last chapter). I would further like to thank Carlos Cordon, Peter Gillies, Jennifer Rietveld, and Mike van Rijswijk for their insights in the Delphi study, with special thanks to Lans Bovenberg. In the past year, I have had the privilege to work together with you, and I hope that our cooperation will continue.

The moral and ethical decision framework

Level 1:
How will it make me feel about myself?
Guided by: publicity, generalization, and mirror test.

Recheck:
Alignment level 1

The moral and ethical question

Level 2:
Is it balanced (for each stakeholder)?
Attention should also be paid to those who will suffer from the decision.

Level 3:
Is it legal?

To all who read this Capstone project, let us try to challenge our personal moral and ethical leadership, our decision-making, and our explanation of those decisions, in order to make a better world for tomorrow and beyond.

REFERENCES

(AFM), A. F. (2019, July 25). *verkenning-ai-verzekeringssector* [Exploration of AI in the insurance sector]. Retrieved from www.afm.nl: https://www.afm.nl/nl-nl/nieuws/2019/jul/verkenning-ai-verzekeringssector

Asimov, I. (1939-1950). I, Robot. Super Science Stories.

Baldini, G., Botterman, M., Neisse, R., & Tallacchini, M. (2016). *Ethical Design in the Internet of Things.*

Bovenberg, P. d. (2019). Where is the love? Over waarde en waarden. *Inaugurele rede 1 november 2019.* Rotterdam.

Brenninkmeijer, A. (2019). *Morele Leiderschap* [Moral Leadership]. Soest: Prometheus.

Bruers, S. (2017). *Morele illusies* [Moral illusions]. Antwerpen: Houtekiet.

Bryant, A. (2011). *Leading Issues in Business Research Methods.* United Kingdom: Good News Digital Books.

Brynjolfsson, E., & McAfee, A. (2016). *The Second Machine Age.* New York: W.W. Norton & Company, Inc.

Buchanan, D., & Huczynski, A. (2017). *Organizational Behaviour.* Harlow: Pearson Education Limited.

Buytendijk, F., Sicular, S., Bretheno, E., & Hare, J. (2019). *AI Ethics: Use 5 Common Guidelines as Your Starting Point.* Gartner.

Clougherty Jones, L., Buytendijk, F., & Hare, J. (2018). *Data Ethics Enables Business Value.* Gartner.

Ellemers, N. (2020, May 22). https://www.naomi-ellemers.nl/188-hoe-maak-je-beslissingen-die-ethisch-juist-zijn-ook-in-tijden-van-crisis [How to make decisions, even in times of crisis]. *Financieel Dagblad*.

Eynikel, J. (2017). *Robot aan het stuur* [A robot at the wheel]. Tielt: LannooCampus.

Faggella, D. (2020, March 14). *AI in law legal practice current applications.* Retrieved from www.emerj.com: https://emerj.com/ai-sector-overviews/ai-in-law-legal-practice-current-applications/

Garz, D. (2009). *Lawrence Kohlberg - An Introduction.* Opladen and Farmington Hills: Barbara Budrich Publishers.

Gillies, P. (2019). *An Experiment: The gods of Cyberanthropology.* Tilburg: TIAS.

Helmer, O., & Dalkey, N. (1951). *The Use of Experts for the Estimation of Bombing Requirements.* Santa Monica: United States Airforce under Project RAND.

Hunt, V., Layton, D., & Prince, S. (2015, January). *Why diversity matters, Mc Kinsey & Company.* Retrieved from www.mckinsey.com: https://www.mckinsey.com/business-functions/organization/our-insights/why-diversity-matters

Johnson, G., Scholes, K., & Whittington, R. (2008). *Exploring Corporate Strategy: Text and Cases (8th edition).* Harlow, Essex: Financial Times Prentice Hall.

Kahneman, D. (2011). *Thinking, fast and slow.* New York: Farrar, Straus and Giroux.

Kalshoven, K., den Hartog, D., & de Hoogh, A. (2011, May). *Ethical Leader Behavior and Big Five Factors of Personality.* Retrieved from researchgate.net: https://www.researchgate.net/publication/226247874_Ethical_Leader_Behavior_and_Big_Five_Factors_of_Personality

Kent, W. (2019, March 26). *Google.com*. Retrieved from www.blog.google: https://www.blog.google/technology/ai/external-advisory-council-help-advance-responsible-development-ai/

Krishna, A. (2020, June 9). *www.ibm.com.* Retrieved from https://www.ibm.com/blogs/policy/facial-recognition-sunset-racial-justice-reforms/?mhsrc=ibmsearch_a&mhq=%20face%20recognition

Luyendijk, J. (2015). *This cannot be true.* Zutphen: Atlas contact.

Maatman, P. (2019). Pensioenfonds en roboadvies: algoritme en aansprakelijkheid [Pension funds and robo-advice: Algorithm and liability]. *Pensioen Magazine*, 6-10.

Melé, D., & Sánchez-Runde, C. (2013, August 7). *Cultural Diversity and Universal Ethics in a Global World (Journal of Business Ethics Volume 116, Issue 4).* Retrieved from Springer.com: https://link.springer.com/article/10.1007/s10551-013-1814-z

Peale, N., & Blanchard, K. (1988). *The power of Ethical Management.* London: Vermilion London.

Peters, J. (2020, June 8). *www.theverge.com/2020/6/8/21284683/ibm-no-longer-general-purpose-facial-recognition-analysis-software.* Retrieved from theverge.com: https://www.theverge.com/2020/6/8/21284683/ibm-no-longer-general-purpose-facial-recognition-analysis-software

Peters, T. (2018, March 8). *www.mt.nl*. Retrieved from mt.nl: https://www.mt.nl/columns/ing-is-crisis-vergeten-nog/551065 [ING has also forgotten the crisis]

Simons, W. (2016, January 16). *MT.nl*. Retrieved from www.mr.nl: https://www.mt.nl/leiderschap/dit-zijn-de-5-grootste-graaiers-euh-meest-besproken-ceos-van-2015/88922 [These are the 5 greediest people—uh, the 5 most talked-about CEOs of 2015]

Sinek, S. (2019). *The Infinite Game.* Business Contact.

Skeepies. (2011, June 15). *https://mens-en-samenleving.infonu.nl/filosofie/75067-wat-is-ehtiek-en-moraal.html#bronnen_en_referenties*. Retrieved from www.mens-en-samenleving.infonu.nl: https://mens-en-samenleving.infonu.nl/filosofie/75067-wat-is-ehtiek-en-moraal.html#bronnen_en_referenties [Humans and society. What is ethics and morality]

Smith, A. (1759). *The Theory of Moral Sentiments.* London.

Smith, A. (1776). *The Wealth of Nations.*

Smith, A. (n.d.). *AdamSmith.org*. Retrieved from Adam Smith Institute: https://www.adamsmith.org/the-theory-of-moral-sentiments

Sorrel, C. (2016, October 13). *Self driving Mercedes will be programmed to sacrifice pedestrians to save the driver.* Retrieved from Fastcompany.com: https://www.fastcompany.com/3064539/self-driving-mercedes-will-be-programmed-to-sacrifice-pedestrians-to-save-the-driver

Tai, E. (2017). Aansprakelijkheid voor robots en algoritmes [Liability for robots and algorithms]. *Nederlands Tijdschrift voor Handelsrecht*, 123-132.

Technologies, E. G. (2018, March 9). *European Commission.* Retrieved from ec.europa.eu: https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf

Thévenot, L., & Boltanski, L. (2006). *On Justification: Economies of Worth.* Princeton University Press.

Tupes, E., & Christal, R. (1961). *Recurrent personality factors based on trait ratings.* Lackland Air Force Base: U.S. Air Force.

Universiteit Utrecht. (n.d.). Retrieved from www.uu.nl: https://www.uu.nl/onderzoek/ethiek-instituut/onderwijs/videomateriaal#utrechtse-stappenplan

van Belleghem, S. (2017). *Customers the day after tomorrow.* Leuven: LannooCampus.

van Dartel, H., & Molewijk, B. (2014). *In gesprek blijven over goede zorg* [Maintaining the dialogue on good care]. Amsterdam: Boom.

van Es, R. (2011). *Professionele ethiek* [Professional ethics]. Deventer: Kluwer.

Van Rijswijk, M., & Teirlinck, M. (2020). *Innoveren vanuit het hart* [Innovation from the heart].

Verbond van Verzekeraars. (2020, June 24). *Mindmap-ethisch-kader-2020.* Retrieved from www.verzekeraars.nl: https://www.verzekeraars.nl/media/7496/mindmap-ethisch-kader-2020.pdf [Mind-map for an ethics framework].