Netspar

SERIES

INDUSTRY

NETSPAR

Challenges of Automated Financial Advice: Definition and Ethical Considerations

Robert Gianni, Minou van der Werf, Lisa Brüggen, Darian Meacham, Jens Hogreve, Thomas Post, Jonas Heller

DESIGN PAPER 236

DESIGN PAPERS are part of the **refereed Industry Paper Series**, which are refereed by the Netspar Editorial Board. Design Papers discuss the design of a component of a pension system or product. A Netspar Design Paper analyzes the objective of a component and the possibilities for improving its efficacy. These papers are easily accessible for industry specialists who are responsible for designing the component being discussed. Authors are allowed to give their personal opinion in a separate section. Design Papers are presented for discussion at Netspar events. Representatives of academic and private sector partners, are invited to these events. Design Papers are published at the Netspar website.

CONTENTS

Affiliations

Robert Gianni – Maastricht University

Minou van der Werf – Maastricht University

Lisa Brüggen – Maastricht University, Netspar

Darian Meacham – Maastricht University

Jens Hogreve – Katholische Universität Eichstätt–Ingolstadt

Thomas Post – Maastricht University

Jonas Heller – Maastricht University

Summary

It is important that individuals are well-informed about their personal financial situation and are able to take action when that situation changes. Automated financial advice could offer new possibilities to increase access to financial advice, and consequently the financial well-being of consumers. Automated financial advice already exists for various financial products, but holistic automated financial advice does not yet exist. In this paper we address two challenges that inhibit its development and implementation: a definition of good automated financial advice, and the ethical issues that come into play. In Chapter 2 we introduce our definition of good automated financial advice, which was created in four iterations with multiple experts. This definition forms a basis for creation of an algorithm for financial advice. In Chapter 3 we dive into the importance of ethics and the ethical issues that play a role in the development of automated financial advice. We discuss the Guidelines for Trustworthy Artificial Intelligence of the European Commission and introduce a practical reflection tool that triggers reflection on the ethical requirements of automated financial advice. In addition, we introduce the AI4 Ethical Financial Services (AI4ES) framework, which helps companies to integrate a viable ethical perspective.

Samenvatting

Het is belangrijk dat consumenten goed geïnformeerd zijn over hun eigen financiële situatie en dat ze in staat zijn om in actie te komen als die situatie verandert. Automatisering zou nieuwe mogelijkheden kunnen bieden om het gebruik van financieel advies te vergroten, met als gevolg een hoger financieel welzijn van consumenten. Geautomatiseerd financieel advies bestaat op dit moment voor verschillende financiële producten, maar holistisch geautomatiseerd financieel advies bestaat nog niet. In dit paper adresseren we twee uitdagingen die de ontwikkeling en implementatie in de weg staan: een definitie van goed geautomatiseerd financieel advies, en de ethische dilemma's die daarbij een rol spelen. In hoofdstuk 2 introduceren we onze definitie, die in vier iteraties is gecreëerd samen met verschillende experts. Deze definitie vormt de basis voor het creëren van een algoritme voor financieel advies. In hoofdstuk 3 bespreken we ethische dilemma's die een rol spelen bij de ontwikkeling van geautomatiseerd financieel advies. We bespreken de 'Guidelines of Trustworthy Artificial Intelligence' van de Europese Commissie en introduceren een praktische reflectietool. Deze tool stimuleert reflectie op ethische vereisten die belangrijk zijn voor geautomatiseerd financieel advies. Daarnaast introduceren we het 'AI4 Ethical Financial Services' (AI4ES) kader, dat bedrijven helpt om op een werkbare manier ethische perspectieven te integreren.

1. Introduction

During their lives, individuals are required to make many financial and other decisions that significantly impact their financial situation and well-being. Buying a house, switching jobs, the arrival of new family members, working more of fewer hours per week, and deciding when to retire are just a few examples of major and often complex events that can significantly alter someone's financial situation and involve related but distinct decisions. Next to these life events, situational changes can also affect a person's financial situation. For example, the rising interest rates that have affected Dutch citizens since November 2021 have significantly increased the prices of energy and groceries, consequently worsening their financial situation (Nibud, 2022; NOS, 2022). Especially people living below or close to the poverty line are hit hard by such price increases. Another example of a situational change that influences a person's financial situation is the current policy change in the Dutch pension landscape, which require people to assume greater responsibility for their financial health in old age. Both types of factors make it ever more important that individuals are well-informed about their personal financial situation, and that they understand the choices they can make to influence their personal situation and the potential consequences of these choices.

Many individuals, however, do not have adequate knowledge or feel insecure when it comes to taking major financial decisions, for example in relation to retirement and old age planning (Eberhardt et al., 2022). According to Nibud, around 40% of persons aged 35 to 55 years have no idea of their financial situation after retirement (Nibud, 2019). In addition, Netspar research has shown that nearly 30% of Dutch households risk having insufficient financial resources after retirement (Knoef et al., 2016). Hence, next to the complexity of many financial decisions, many people often too little knowledge to make these choices by themselves. Therefore, many individuals would probably benefit from financial advice regarding their financial decisions. The AFM also argues that financial well-being often begins with sound financial advice, due to the complexity of financial products and the related decisions (AFM, 2018). However, half of the Dutch population indicates that the cost of financial advice constitutes a barrier for them to seek such advice, and one third indicates not being sure as to whether a financial advisor would work in their best interest (Nibud, 2017). It is especially persons with a below-average income, a part of the population that would arguably benefit the most, who indicate that they search things out themselves to avoid the high costs of an advisor.

Technological developments can offer possibilities to increase access to and acceptance of financial advice, and consequently to raise the financial well-being of individuals. One such possibility is automated financial advice (sometimes referred to as robo-advice[1]). We define automated financial advice as an algorithm, which incorporates the socio-demographic characteristics, preferences, financial behaviors, and goals of individuals to assess their current financial situation and, based on that assessment, recommends specific financial and other actions. Automated financial advice already exists for various financial products, such as auto and health insurance and portfolio management. However, a study on more holistic automated financial advice does not yet exist in the market. That is the focus of this paper.

Automated financial advice: what are we talking about?

When we speak of automated financial advice, we mean a machine-generated recommendation for an individual or household in relation to a specific financial action or choice, on the basis of personal situation and preferences. We are not talking about product advice that falls under the Financial Supervision Act (*Wet financieel toezicht, Wft*). In the context of the new pension arrangement in the Netherlands, for example, automated financial advice could be used to assist participants in choosing a fixed versus a variable annuity; to decide on whether to take out a lump-sum amount at retirement; or, even more broadly, to assess their financial situation and get insight and advice on whether it is sufficient for the goals that participants have for retirement.

We see automated financial advice as a step beyond choice guidance. In our view, choice guidance entails guiding individuals or households, for example by tailoring options. This can include calculating the specific consequences of a choice for the financial situation; only presenting options that apply for the personal situation; or ranking the choice options by how appropriate they are for one's financial situation and preferences. The significant difference for us is that choice guidance does not give specific recommendations.

N.B. Even though in this paper we focus on automated financial advice, most conclusions also apply to choice guidance.

The advantages of automated financial advice include the possibility of reducing costs, increasing the quantitative level of analysis and potential solutions, and improving the objectivity and overall quality of advice. This, in combination with the fact that automated services can be readily accessed at home and at any hour, could significantly increase access to and uptake of financial advice. This increase would be an advantage for both end-users and providers. However, despite these advantages

---

1 As robo-advice often merely focuses on investments, we use the term automated financial advice in this paper.

and the technical possibility to create automated financial advice, there are also significant challenges that prevent the development and implementation of holistic automated financial advice. For example, in a sensitive domain like personal finance, it is important that digital tools such as automated financial advice work demonstrably in people's best interest. However, a sufficiently comprehensive identification of what is in "people's best interest" is lacking. Therefore, in Chapter 2 we introduce a definition for "good" automated financial advice, created with input from academic and industry experts.

In Chapter 3, we attend to the ethical implications that are associated with automated processes. Discrimination, paternalism, and lack of transparency in terms of the underlying objectives of automated financial advice are negative aspects that can have severe repercussions, which can induce harm for certain groups and affect trust in the algorithmic translation of financial advice (Coeckelbergh 2019, Crawford 2021). Thus, when automating human processes, special attention to unintended (as well as intended) effects is required. In this chapter, we also introduce a practical tool that aims to trigger reflection on the ethical requirements of automated financial advice that are important, plus a framework that is meant to help companies to integrate a viable ethical perspective.

2. Definition

In a sensitive domain such as personal finance, it is important that advice is demonstrably in the best interest of the individuals or households receiving the advice. If this is not the case, it may end up influencing their financial and even overall well-being in a negative way, since the impact of an individual's or family's financial situation stretches across all aspects of life (Chapman & Freak, 2013; Drentea, 2000; Drentea & Lavrakas, 2000; Dunn, Gilbert, & Wilson, 2011; Lane, 2016; Social Science and Parliamentary Affairs Team, 2010; Van Dijk, 2016). However, to date it is unclear 'good' financial advice should be defined, because the evidence base for what is in people's best interest is lacking. While there have been some suggestions on what might constitute good financial advice, the literature agrees primarily that there is a need to develop benchmarks for good financial advice. This is especially relevant with the current developments towards automated financial advice, as this benchmark should form the basis for the financial advice algorithm.

We consequently have developed a definition of good automated financial advice[2]. To come to this definition, we first organized an interactive ideation workshop with experts from APG, AFM, Ortec Finance, Nibud, KVCP, and Maastricht University, in which we discussed the definition. Based on the input, we derived six key elements of good financial advice:

1. Increased well-being: Financial advice is good if it leaves consumers happy and positively increases their subjective and objective financial well-being.
2. Alignment with short- and long-term goals: Financial advice is good when it is directed at the achievement of the individual person's goals. Good financial advice should balance the trade-off between short-term and long-term personal goals.
3. Ease of understanding: Financial advice is good when it is easy to understand. The individual should not need to put too much effort or time into understanding the advice. Good financial advice should be understandable for all.
4. Incorporation of individual preferences: Financial advice is good when individual preferences are incorporated into the advice that is given. The options that financial advice provides to reach financial and other goals should be within the limits of personal preferences (e.g., risk preference / time preference).

---

2 Even though the definition focuses on good *automated* financial advice, it applies almost everything to all human financial advice.

5. Ease of implementation: Financial advice is good when individuals take up the advice and implement it. Good financial advice should focus on which practical steps individuals should take to achieve their goals.
6. Transparency: Financial advice is good when it is transparent. Consumers should be able to trust the advice, and this requires it to be transparent.

We used this input on what constitutes good financial advice to draft a first version of the definition. To validate whether this definition truly encompasses what experts see as good financial advice, we presented the definition to professionals (academic experts, policy makers, pension providers) at a digital round table organized by Netspar in January 2022. In break-out sessions, participants discussed their understanding of "good financial advice". In a follow-up survey, we asked them to evaluate our working definition. The main feedback from the sessions and the survey was that the current definition:
– is too complicated;
– only talks about providing strategies, whereas it should also incorporate the positive and negative consequences of choices;
– should include that the advice be independent;
– should include that the foundation of the advice is built upon accurate and complete data.

The feedback was used to further refine the definition. In a third session (March 2022) with experts – from APG, AFM, a.s.r., Ortec Finance, the Catholic University of Eichstätt-Ingolstadt, and Maastricht University – this new definition was evaluated, leading to some further minor adjustments.

Good automated financial advice: a definition

Basis. Good automated financial advice is based on sufficient, relevant, and accurate data on the financial situation and personal preferences (i.e., risk profile, life events, goals) of the individual. Next to that, the algorithm should be objective and transparent in the way it reaches the advice.
Goal. Given this basis, good automated financial advice optimizes the person's subjective and objective financial and other well-being.
Outcome. Good automated financial advice gives individuals insight into their financial situation and the positive and negative consequences and trade-offs of their possible choices. It gives them concrete and applicable actions that help them to improve their financial and other well-being in their specific situation.
Communication. Good automated financial advice is understandable for everyone.

As a final iteration, the definition was presented during a Netspar conference to around fifty Netspar partners and research fellows. This led us to this final definition of good automated financial advice, which is divided into four components. It defines (i) what goal good automated financial advice should have, (ii) what information the algorithm should be based on and how it should operate, (iii) what the outcome of good financial advice should look like, and (iv) how it should be communicated.

3. Ethics

The definition of good automated financial advice forms a necessary basis for creating an algorithm for financial advice. However, it takes more than just a definition. In this chapter, we look at ethical issues related to automated financial advice. We purposely choose to take an ethical rather than a legal perspective. While laws and regulations are of course fundamental grounds, they are not always sufficient for addressing certain issues that could (intentionally or not) affect people's financial and general well-being, and hence also societal well-being. Legal compliance is automatically a key aspect when processing sensitive data (Carrillo 2020), but there is often less attention to the ethical considerations related to automated decision-making or decision-support processes. This is due to the fuzziness of many ethical questions, the subsequent difficulty in operationalizing ethical frameworks, and the shifting public attitudes towards automated decision-making in various spheres (Aysolmaz, Muller & Meacham 2023). Ethical considerations are nonetheless extremely important due to the potentially high impact that automated decision-making, and automated financial advice in particular, can have on the lives of individuals, households, and communities.

> **Ethics**
>
> Ethics can help evaluate potential resistance and preferred aspects of technological innovations that are not subject to dedicated regulation. Ethical frameworks take a broader perspective and focus on novelties and changes to improve societal well-being. If we apply an ethical perspective to potential issues arising from automated financial advice, we can address aspects that might jeopardize the adoption of automated services beforehand and increase overall trust (Bedué and Fritzsche 2021, Glickson and Woolley 2020).

Although individuals may think that decisions made by automated systems are more objective and not subject to bias[3] like decisions made by humans, the data and algorithm are informed by specific norms, values, and worldviews of the people who collect or work with the data or algorithm, as well as by the extent to which the data sets are representative. Consequently, when automated systems are used to determine or advise on outcomes for human subjects, it is important to attend also to issues concerning accountability, security, privacy, transparency, accuracy, bias, value sensitivity, and appropriateness associated with their use.

---

3   With bias we mean the action of supporting or opposing a particular person or thing in an unfair way, by allowing personal opinions to influence judgment.

3.1 Ethical issues

In the previous paragraph, we briefly touched upon the main ethical issues that arise from the introduction of automated financial advice systems. In this section we elaborate in more detail on those issues. When adopting automated financial advice, we can identify the following aspects that can raise concerns: autonomy and control, discrimination, harm, and transparency lack of transparency (opacity).

Autonomy and control

A first issue is how to maintain control over increasingly autonomous Artificial Intelligence (AI)[4]. The level of autonomy of the automated system versus human control is a heavily debated topic in the academic, policy, and public debate. Reducing the autonomy of automated systems through increased human oversight and intervention may reduce their potential for efficiency gains, but it may also help increase public acceptance. Increasing the scope for automated decision-making in sensitive spheres of human activity (e.g., health, finance, security) will require rethinking well-established norms as well as such issues as the moral and legal apparatus concerning accountability and responsibility. Imagine, for example, that a system with a high degree of autonomy, that learns from the external environment of the potential investor, were to decide to deviate significantly from the initial code, picking a different option than initially foreseen. In such a case, who or what entity would be accountable in the event of damage, such as a severe financial loss?

A second issue is nudging by AI. Automated systems can nudge individuals towards solutions preferred by the service provider. Advice that is generated by an algorithm is often perceived as being more objective, but the algorithm of the machine simply expresses a design that has been previously decided upon by the developers. Individuals who receive suggestions from automated systems may not be aware of this mechanism, possibly weakening the critical attention that they give to the quality of the advice and to any potential conflict of interest behind it. Hence, using AI for financial advice may strengthen a process of nudging that people have even fewer means of resisting against.

---

4 AI is an umbrella term for a variety of different approaches that design technical devices to perform operations similar to human cognition, but faster and with an estimated higher accuracy (Bawack et al., 2021). Although there is still discussion as to whether the use of automated systems in the field of financial advice is based on simple computational or AI-based technologies, it is plausible to predict that the role of AI in the field of financial advice is set to grow significantly, as the adoption of AI is growing in all fields. Moreover, ethics guidelines for AI can also be informative for automated systems that do not yet involve AI.

Difficulties inherent in finding the right balance between the development and exploitation of technological possibilities on the one hand, and respect for human values and dignity on the other, will likely arise in the context of automated financial advice. Subsequently, the level of autonomy we consider appropriate in order to maintain a balance between automated financial advice that can improve financial and other well-being, and the necessary autonomy of the humans interacting with AI, is a core ethical question in the development of automated financial advice systems.

Discrimination
Using data and algorithms may result in discrimination, as both are created by biased (consciously or unconsciously) humans. When data and algorithms are used in decision-making processes, implicit biases that the data and/or developer has can be inadvertently passed onto the algorithm and the recommendations it makes. It could, for example, be that the data do not include certain groups (such as minorities), for which the optimal financial choice might differ from the individuals represented in the data. It is probably impossible to exclude biases altogether. However, if those biases are not made explicit, they could turn into discriminatory practices, such as infringing upon fundamental rights such as the autonomy of the individual and the respect for human dignity.

Harm
Automated financial advice is not immune to doubts about its potential harm. In fact, the risks mentioned above are increased by the potential harm that could be inflicted on human subjects who make use of it. The harm of automated financial advice would manifest itself in financial choices that do not correspond with the profile of the end-user, and with the chance of decreasing rather than increasing someone's subjective and objective financial and other well-being.

Opacity
The abovementioned risks also all connect to the opacity of the process. What data are collected, and what are not? What is the combination of the different sets of information and according to which criteria does the AI make a decision? What is the underlying objective of the analysis process? Too much opacity generates skepticism and distrust, as the functioning of AI can be unclear and generate doubts about the data being collected and how the AI makes decisions.

The risks of these ethical issues can emerge or increase when the automated decision-making process is not adequately transparent and there is a lack of accountability in the overall implementation. The question then is how to prevent these risks from happening and/or how to mitigate them. In the next section, we will talk about how ethical frameworks and guidelines can help us prevent or mitigate risks related to the implementation of AI-based systems.

3.2 Ethical frameworks and guidelines

There is a growing consensus that designing and implementing sufficient regulatory and ethical safeguards for the use of AI can significantly improve the societal acceptance and adoption of these technologies (Floridi et al., 2018). There is also growing consensus about the difficulty of implementing often high-level ethical frameworks in concrete contexts (Bleher & Braun, 2023). Therefore, to address ethical issues in automated financial advice, we rely on ethical frameworks that have been developed in the context of AI. According to the AI Index Report, 117 documents addressing AI principles were published between 2015 and 2020. Although the number of ethical guidelines is significant and is increasing at a fast pace, analyses have shown that almost all documents tend to adopt remarkably similar criteria (Hagendorff, 2020; Jobin et al., 2019).

Hagendorff (2020) has, for example, shown that aspects of fairness, privacy, and accountability are deemed necessary in nearly 80% of AI ethical guidelines to design an ethically sound AI. However, not all documents exude the same legitimacy or normative force. For the purpose of this paper, we follow the indications provided in the "Ethics Guidelines for Trustworthy AI" published by the European Commission (EC, 2019), for four reasons:

1. In drafting the document, 52 experts from different disciplinary fields and backgrounds deliberated for months, integrating the outcomes of consultations in civil society.

2. This document builds on and develops humanitarian analyses and policies that have been in place for quite some time, and enriches it with a wealth of human rights experience.

3. The guidelines make the values behind the requirements explicit, plus their interrelated functionality in moving from more abstract to more pragmatic levels.

4. The European Commission is the main public funding organization in Europe, making these guidelines an important shared reference point at European level.

Table 1: Examples of value-based principles for trustworthy AI

Organization	Value-based principles for trustworthy AI
EU high-level expert group on the ethics of AI	1. Respect for human autonomy 2. Prevention of harm 3. Fairness 4. Explicability
OECD	1. Sustainability 2. Fairness 3. Transparency 4. Safety 5. Accountability
Google	1. Socially beneficial 2. Avoid creating or reinforcing bias 3. Safe 4. Accountable to people
Microsoft	1. Fairness 2. Reliability 3. Security 4. Privacy 5. Inclusiveness 6. Transparency

3.3 Ethical guidelines for trustworthy AI

Values

An important aspect for a trustworthy AI is the driving values and principles that inspire and justify the selection of specific normative tools. The basic values at the heart of the 'Ethical Guidelines for Trustworthy AI' are derived from the fundamental values of the European Union and the relevant foundation documents. The four values identified by the High Level Expert Group are the following:

I.   *Respect for human dignity*

II.   *Freedom of the individual*

III. *Respect for democracy, justice, and the rule of law*

IV.   *Equality, non-discrimination, and solidarity*

These values are abstract and do not necessarily address the field of financial advice. Therefore, it is important to translate them into principles that will subsequently lead to the requirements for trustworthy automated financial advice.

Principles

Principles provide more concrete guidance for understanding the normative rationale behind a specific approach. These principles are:

Figure 1: Summary of the values, principles and requirements
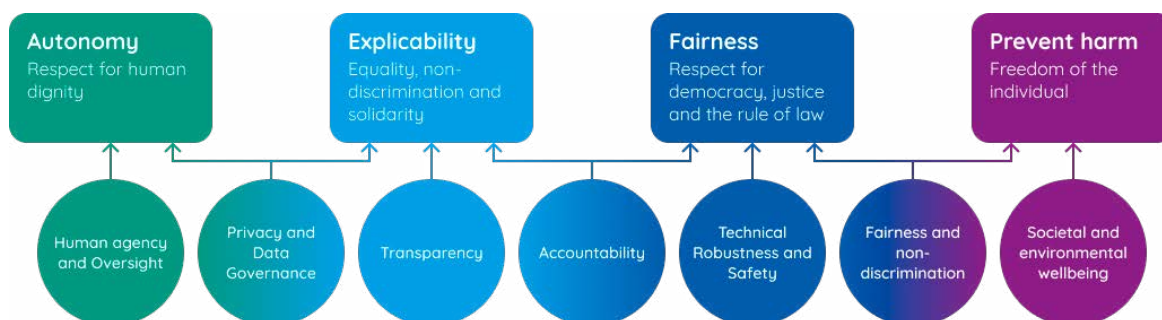
a) Human autonomy

Respect for human autonomy, meaning that AI systems should be designed, and most of all implemented, with respect for the necessary autonomy of human beings interacting with them. Furthermore, the AI should be modelled such as to augment, complement, and empower human capacities and not endanger or diminish them.

b) Prevention of harm

Automated financial advice systems should be designed, deployed, and implemented according to the highest levels of technical robustness. They should also be monitored to prevent exacerbation of negative impacts due to asymmetries of power, information, and other forms of vulnerability.

c) Fairness

The development, deployment, and use of AI systems must be fair. This means that designers and organizations implementing AI-based systems should ensure a fair distribution of costs and benefits, but also that the procedures are established on the basis of unbiased and equal criteria.

d) Explicability

Explicability is crucial for building and maintaining users' trust in AI systems. Explicability implies that the most important parts of the process are communicated on a timely basis and are intelligible to those directly and indirectly affected.

In the next section, we translate these values and principles into concrete requirements, and analyze how they apply for the adoption of automated financial advice.

Requirements

1. *Human Agency and Oversight*
    Human beings must be able to maintain a certain level of control over automated decision-making processes. One can imagine that the optimal system is a multi-level approach to human oversight and control depending on the level of risk and extension in time: the greater the risk and the longer the investment, the smaller the role of AI. Even though AI could help to increase objectivity and to diminish potential conflict of interest, we believe that, in the current state of AI development, it is more prudent to take a risk-adverse approach that tends toward greater human control in decision-making processes, correlated to risk level and durational impact of the decision-making. Accordingly, service providers may want to choose the governance system in terms of Human in Control (HIC), human in the loop (HITL), or human on the loop (HOTL). In the case of automated financial advice, human control should address service providers and end-users, by ensuring that service providers have complete control of the process, and that end-users are free to choose the extent of AI processing.

2. *Technical Robustness and Safety*
    The automated financial advice system should be designed and deployed according to high measures of security against breaches or intentional and unintentional abuse of the end-users' personal data. The system should be operated with the possibility that a human operator can interrupt the process to verify its reliability. Automated financial advice should be constructed such as to guarantee that the system does not deviate from its foreseen functions, and the security measures should be proportional to the level of risk that a breach can generate.

3. *Privacy and Data Governance*
    Automated financial advice systems should be built according to privacy-by-design[5] and deployed through constant monitoring of the additional data generated by the system and subsequent interaction with the end-users. For instance, automated financial advice will maintain a data exchange relationship with end-users about financial decisions and changes in their financial situation. This requires data governance plans that explain the entire process of data collection, use,

---

5   *http://jpaulgibson.synology.me/ETHICS4EU-Brick-SmartPills-TeacherWebSite/SecondaryMaterial/pdfs/CavoukianETAL09.pdf*

storage, and accessibility (who can access the data, what data, for what reasons, etc.).

4. *Transparency*
Transparency involves the need to know whether and how an end-user interacts with an AI-based system, plus the possibility of understanding how and why certain choices for advice are made by the organization and the AI. Transparency is thus an umbrella term that can be divided into three sub-categories:

   a. *Traceability:* Documentation on data sets and processes (e.g., algorithms and criteria used) needs to be maintained. This makes it possible to understand the origins of any errors or unintended consequences and to improve the robustness of the process.

   b. *Explainability:* It should be possible to explain the underlying rationale behind financial decisions that are selected by the AI for the end-user.

   c. *Communication:* Organizations should inform end-users about the use of automated systems and specify the modalities, advantages, and possible limitations compared to human financial advice. Organizations should also guarantee that end-users have the possibility to not use automated financial advice systems.

5. *Fairness and non-discrimination*
Automated financial advice systems should be designed on the basis of data and processes that are non-discriminatory and inclusive. In addition, organizations must try to avoid unfair competition in agreement with other organizations. Finally, automated financial advice systems need to be designed with a focus on their accessibility, regardless of factors such as age, education, digital skills, or disabilities (Art. 42 Public Procurement Directive "Design for all"). Whenever possible, and in the ways available to the organization, it is advisable that the fairness of the system is tested with internal and external stakeholders, thereby reinforcing the validation process (ethics advisors, sector associations, etc.).

6. *Accountability*
Organizations should be prepared to have some of their datasets available for audit. This aspect can significantly lead to greater trustworthiness and thus increase their adoption by end-users.

7. *Societal and environmental well-being*

Automated financial advice should not lead to a deteriorating function for the employees within the organization. Additionally, access should be guaranteed in an equal manner to all potential end-users without prejudice. Service providers should encourage investments that have a positive impact on environmental well-being, as well as those directed at addressing societal challenges.

3.4 Practical guide: reflection tool

In the previous section we moved the discussion from highly abstract values to more concrete, but sometimes still abstract, requirements. Therefore, we have created a reflection tool (see Appendix I) that is inspired by the European Commission guidelines for Trustworthy AI. This tool aims at translating the abovementioned requirements into concrete questions that can guide practitioners in their understanding of ethics and in its implementation in the design of AI.

The questions listed in the reflection tool are only examples of how the different requirements can be translated into practical action. However, practitioners and designers may well wish to extend the set of questions to tailor the requirements to the capacity of their organizations. This reflection tool is only intended to be a non-exhaustive pathway to touch on some of the aspects and to offer food for thought. It can be used at any stage of the AI development, but it is highly recommended to be used in the design phase and prior to implementation of the automated financial advice [6].

Human Agency and Oversight
- Can you as a service provider access, monitor, and possibly interrupt the AI process without causing undue damage (e.g., breaking the system, losing financial gains)?
- Have you anticipated different levels of autonomy of the AI within the automated financial advice system?
- How many levels have you planned to introduce?
- Do they vary according to the extent of the risk?
- Is it possible to withdraw from what the AI is proposing, in favor of human interaction?
- Is there a human-in-the-loop who is able to monitor the accuracy, robustness, compliance, and fairness of the AI system?

6 Next to the implementation phase, it is also important to keep in mind that an evaluation of the methods employed to implement the requirements, as well as to reporting and justifying changes to the implementation process, should occur on an ongoing basis.

Figure 2: Example questions of the reflection tool that we created to guide practitioners in their understanding of ethics and its implementation in the design of AI.
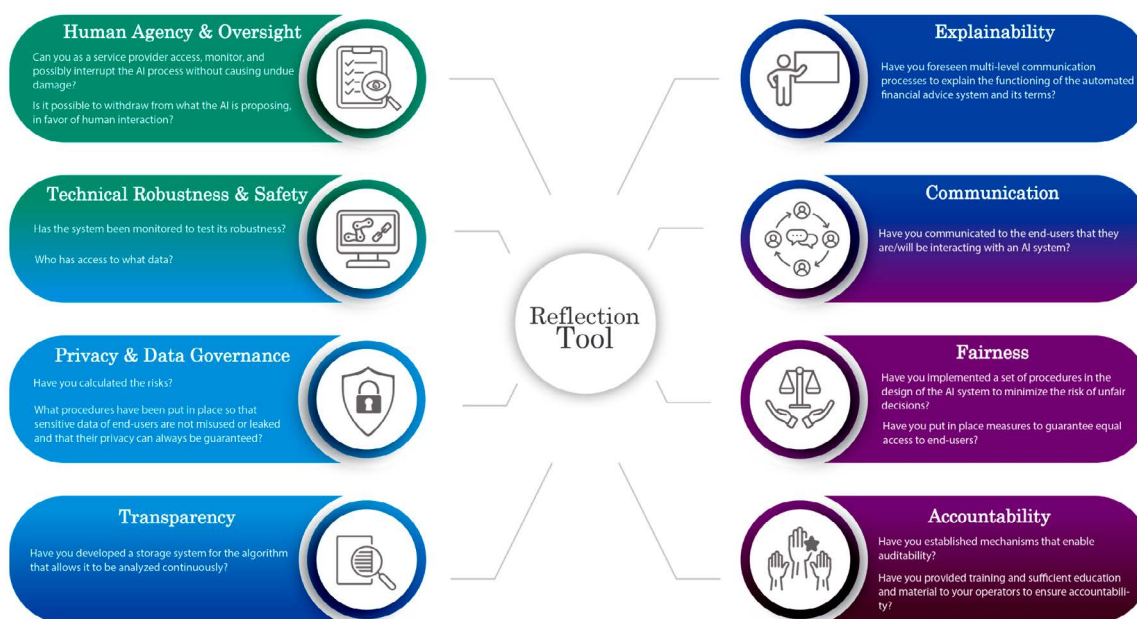
– At what stage of the implementation and how frequently are robustness, compliance, and fairness checked?

Technical Robustness and Safety
– What kind of tests has the system passed?
– Has the system been monitored to test its robustness?
– Who has access to what data?
– Have you put in place a mitigation plan in case of system failure or breakdown?

Privacy and Data Governance
– Have you put in place procedures that address this aspect?
– Have you calculated the risks?
– Have you put in place measures to protect against risks?
– Have you put in place a mitigation plan?
– What procedures have been put in place so that sensitive data of end-users are not misused or leaked and that their privacy can always be guaranteed? What mitigation plan has been designed in case of a breach?

Transparency
- Traceability:
  - Have you developed a storage system for the algorithm that allows it to be analyzed continuously? Examples of this can include the following, without being exhaustive:
  - The methods for training the algorithm, including the input data collected and selected and how they were identified.
  - The methods for testing and validating the algorithm.
  - The outcomes of the algorithmic system.
- Explainability:
  - Have you foreseen multi-level communication processes to explain the functioning of the automated financial advice system and its terms?
  - Is the automated financial advice system also accessible for end-users without great expertise or with limited digital skills?
- Communication:
  - Have you communicated to the end-users that they are/will be interacting with an AI system?
  - Have you clarified aspects related to the advantages and potential limitations of using an AI system?
  - Are you able to inform the end-users about the criteria that inform the AI suggested advice?

- *Fairness*
  - Have you implemented a set of procedures in the design of the AI system to minimize the risk of unfair decisions?
  - Have you put in place measures to guarantee equal access to end-users?
  - Have you implemented a process of monitoring and potential identification of biases or unfair processes?
  - Have you carried out an impact assessment in terms of the accessibility of the system?
- *Accountability*
  - Have you established mechanisms that enable auditability?
  - Have you carried out a risk assessment of the AI system?
  - Have you provided training and sufficient education and material to your operators to ensure accountability?
  - Have you foreseen redress mechanisms/procedures in case of adverse impact?

3.5 AI4ES: incorporating ethics in your organization

Responses to all questions in the reflection tool are seldom immediately available to developers or responsible figures within financial advice organizations. Accordingly, organizations are encouraged to ensure that diverse teams work on such challenges. In order to help organizations to do this, we have created the AI4 Ethical Financial Services (AI4ES) framework, a multilevel approach for including ethics in financial services (see Figure 2). It consists of four levels according to the variety and effective inclusion of different perspectives in the decision-making process. The various levels are meant to help service companies with diverse needs, resources, and scope of application in integrating a viable ethical perspective without losing sight of the potential impact on their automated financial advice service. The levels of ethical impact vary according to the degree of inclusiveness of external actors as well as their influence in the process.

Level 1: Charter of Ethical Requirements

The idea of Level 1 is that the design team should try to integrate ethical guidelines and requirements into technical design and construction processes, ensuring that the design and production processes are as robust as possible. This can be done, for example, by including a charter in the AI design phase that integrates a series of recommendations or requirements that are drawn from the ethical principles proposed by the European Committee, as described above. These ethical principles should function as guidance for further reflection – for which the reflection tool that we describe can be used – according to the features that characterize the ecosystem[7] where the AI is going to be designed and implemented. Additionally, one might think of ways of integrating ethical considerations into the development of AI and

---

7   'An ecosystem is a system of resource-integrating actors connected by shared institutional arrangements and mutual value creation through service exchange' (Vargo and Lusch 2016, p. 11). It is the environment where the design, deployment, implementation, and consumption of the AI service take place. It includes the infrastructure, actors, behaviors, and specific goals of AI, as well as the political, social, and cultural environment in which a service is provided, as these conditions can vary and influence the optimal delivery of a service. Following the indications named as examples by the European Commission, we also consider that AI ecosystems are characterized by three main stakeholder groups, namely designers, deployers, and end-users. *Designers* are the technical experts "who research, design and/or develop AI systems"; *deployers* are "public or private organizations that use AI systems within their business processes and that offer products and services to others"; and *end-users* are those individuals "engaging with the AI system, directly or indirectly".

emotions in services, in order to increase awareness. For example, by introducing a Hippocratic oath[8] for engineers and designers, like that taken by doctors, to make designers more aware of their responsibility and the impact of their choices (Williams 2018). A drawback of this level is that designers and engineers may find it difficult to understand and correctly apply these ethical requirements, possibly leading to the ethical requirements not being embedded sufficiently or correctly into the AI-based service design process.

Level 2: Internal diversity
The aim of the second level is to foster the integration of a broader consideration of ethical issues that arise in the development of services using AI, as well as manners for addressing them. This is done through diversifying the composition of the relevant teams as much as possible. Diversity should be understood as diversity of perspectives that can be linked to forms of diversity such as gender, culture, and socio-economic background or diversity on other dimensions. However, diversity is not only a question of broadening demographics, but also the types of expertise that developers of AI draw upon to address ethical issues. For instance, organizations could integrate social scientists or psychologists, who could work closely with data scientists and help them consider broader ethical aspects. It is also important to consider diversity among the social scientists, psychologists, or ethicists brought into these groups. If deployers are responsible for the hiring processes, designers should be aware of this need for diversity and help the deployers with adequate suggestions. A drawback of this level is that internal diversity might not be sufficiently independent and neutral to provide a truly diverse perspective. Alignment to a company's objectives as well as its internal dynamics can limit the broadness in the design of AI.

Level 3: External diversity
The third level involves suggestions from and cooperation with external actors, by introducing regular external audits by experts, who evaluate legal compliance, technical robustness, and ethical appropriateness from a more objective stance. The

---

8   An example of this could be: "*As someone who shapes the lives of others, I promise to: Care genuinely about their success; Understand their intentions, goals, and values as completely as possible; Align my projects and actions with their intentions, goals, and values; Respect their dignity, attention, and freedom, and never use their own weaknesses against them; Measure the full effect of my projects on their lives, and not just those effects that are important to me; Communicate clearly, honestly, and frequently my intentions and methods; and Promote their ability to direct their own lives by encouraging reflection on their own values, goals, and intentions*".

Figure 3: Visual representation of the AI4 Ethical Financial Services (AI4ES) framework
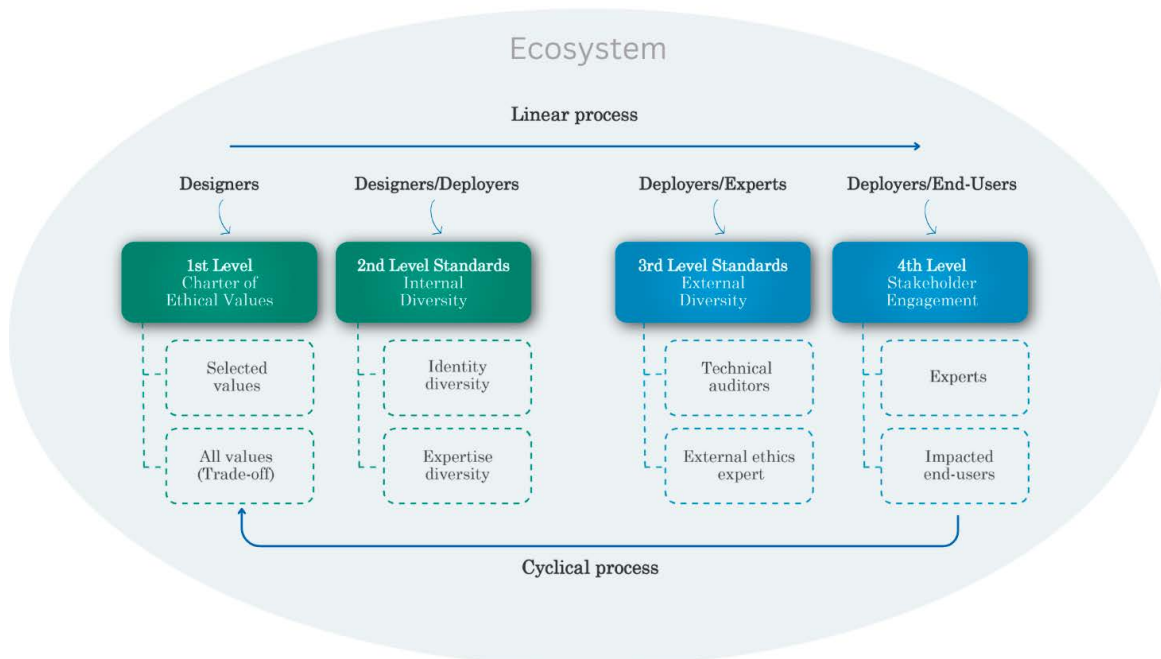
advantage of using external experts is that they are independent and neutral and can therefore provide a more critical and diverse set of reflection and input. Just as in the other levels, the group of experts should be diverse. Service organizations can choose the format (an advisory board or recruitment on demand), frequency, and depth of such feedback, considering the trade-off between quality and cost. A drawback of this level is that it is resource-intensive.

Level 4: Stakeholder engagement
The fourth level aims to increase the contribution of external actors, by stakeholder engagement (including end-users) in the process. This level has the great advantage that it increases the diversity of knowledge. Stakeholder engagement can be orchestrated at various stages in the service development process and in a variety of ways, including surveys and interviews to measure stakeholders' perceptions, or interactive workshops in which the AI is co-created with designers, deployers, experts, and end-users. Feedback provided by end-users can reveal whether the direction taken by the company is the right one or whether there is room for improvement. End-users represent a key demographic element for efforts to increase the level of acceptability and acceptance of a product. For one thing, they can increase the legitimacy of a specific design of the AI, thereby limiting potential skepticism and rejection because of ethical reasons. Engaging a wide range of stakeholders is currently the most common

procedure to increase the democratization of research and innovation. In addition, including a wide range of end-users can increase the actual acceptance of a product or process, as they can contribute to the evaluation of different aspects, providing suggestions on how to improve the product or the process. Accordingly, while more challenging, stakeholder engagement can bring significant benefits from an ethical but also economic point of view.

How to understand the levels
The four levels graduate in terms of involvement, costs, and complexity, but they are complementary and not exclusive. While the levels build upon each other, they can also be seen as cyclical. For instance, service providers could establish a stakeholder engagement process that is managed internally, incorporating the seven requirements of the AI charter. In an ideal scenario, end-users cyclically scrutinize the adopted ethical measures. Evaluation should be start of the design phase, to influence the process effectively. As AI services as well as the ecosystems evolve at a fast pace, it is important to maintain a dynamic and stable process of interaction between design-ers, deployers, and end-users.

4. Conclusion

In this paper we have introduced a definition of good automated financial advice, and we have discussed the ethical issues that play a role within automated financial advice. Our definition of good automated financial advice is divided into four components and defines (i) which goal good automated financial advice should have, (ii) on which information the algorithm should be based, and how it should operate, (iii) what the outcome of good financial advice should look like, and (iv) how it should be communicated. This definition could form the basis for an algorithm for automated financial advice. We have also discussed the ethical issues that might play a role in the development of automated financial advice and have introduced a practical reflection tool and a framework, both of which help companies to integrate a viable ethical perspective. Applying an ethical perspective to the potential issues arising from automated financial advice prevents aspects that jeopardize the adoption of automated services from not being addressed beforehand. In this paper, we have described the ethical issues from a theoretical perspective, based on existing literature and frameworks.

Limitations and future research
This paper is subject to several limitations. The first limitation is that we have not validated our definition of good automated financial advice with the intended end-users of the service. In addition, we have not incorporated concerns about broader societal or environmental impact into the definition, beyond the concerns that may be held by specific individuals or households. We also have not taken into consideration environmental concerns linked to the computing power required for increasingly complex automated decision-making systems (Monserrate, 2022). These would all be items to investigate in future research. A further next step in the research would be to carry out qualitative research with experts in the field concerning their experiences with these ethical issues. A second further step would be to conduct a comparative analysis of the reflection tool elaborated in this paper with other existing data-ethics tools, for example the ODI data-ethics canvas[9] or the Data Ethics Decision Aid (DEDA).[10]

---

9   *https://theodi.org/article/the-data-ethics-canvas-2021/*
10 *https://dataschool.nl/en/deda/*

References

AFM (2018). *Visie op roboadvies: Kansen, zorgplicht en aandachtspunten.* Retrieved from https://www.afm.nl/~/profmedia/files/onderwerpen/roboadvies-sav/visie-roboadvies.pdf?la=nl-nl

Anderson, E. (2006). The epistemology of democracy. *A Journal of Social Epistemology, 3*(1), 8-22.

Aysolmaz, B., Müller, R., and Meacham, D. (2023). "The public perceptions of algorithmic decision-making systems: Results from a large-scale survey", Telematics and Informatics, V. 79, https://doi.org/10.1016/j.tele.2023.101954

Bedué, P. & Albrecht F. (2021). Can we trust AI? An empirical investigation of trust requirements and guide to successful AI adoption. *Journal of Enterprise Information Management, 35*(2), 530-549. doi: 10.1108/JEIM-06-2020-0233

Bleher, H. & Braun, M. (2023). Reflections on Putting AI Ethics into Practice: How Three AI Ethics Approaches Conceptualize Theory and Practice. *Science and Engineering Ethics,* 29, 21. https://doi.org/10.1007/s11948-023-00443-3

Carrillo, M. R. (2020). Artificial intelligence: From ethics to law. *Telecommunications Policy, 44*(6). doi: 10.1016/j.telpol.2020.101937

Chapman, S. J. & Freak, M. (2013). Personal finance. In S. J. Chapman & M. Freak (Eds.). *New concepts in commerce* (pp. 40–73). Brisbane, Australia: John Wiley and Sons.

Coeckelbergh, M. (2019). Artificial Intelligence: Some Ethical Issues and Regulatory Challenges. *Technology and Regulation,* 31-34. doi: 10.26116/techreg.2019.003

Das, S. K., Aritra D., Akash P., & Nabamita Banerjee R. (2015). Applications of Artificial Intelligence in Machine Learning: Review and Prospect. *International Journal of Computer Applications, 115,* 31-41.

Drentea, P. (2000). Age, debt, and anxiety. *Journal of Health and Social Behavior, 41,* 437–450.

Drentea, P. & Lavrakas, P. J. (2000). Over the limit: The association among health, race and debt. *Social Science & Medicine, 50,* 517–529.

Dunn, E. W., Gilbert, D. T., & Wilson, T. D. (2011). If money doesn't make you happy, then you probably aren't spending it right. *Journal of Consumer Psychology, 21,* 115–125.

Eberhardt, W., Post, T., Hoet, C., & Bruggen, E. (2022). Exploring the first steps of retirement engagement: a conceptual model and field evidence. *Journal of Service Management, 33*(6), 1-26. doi: 10.1108/JOSM-11-2020-0402

European Commission, Directorate-General for Communications Networks, Content and Technology. (2019). *Ethics guidelines for trustworthy AI.* Retrieved from https://digital- strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai

Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., & Dignum, V. (2018). AI4People: An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines, 28*(4), 689-707.

Gianni, R., Lehtinen, S., & Nieminen, M. (2022). Governance of Responsible AI: From Ethical Guidelines to Cooperative Policies. *Frontiers of Computer Science, 4.* doi: 10.3389/fcomp.2022.873437.

Glickson, E. & Williams Woolley, A. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals, 14*(2), 627-660. doi: 10.5465/annals.2018.0057.

Haas, L. & Gießler, S. (2020). *In the realm of paper tigers – exploring the failings of AI ethics guidelines.* Retrieved from https://algorithmwatch.org/en/ai-ethics-guidelines-inventory-upgrade- 2020/

Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds & Machines, 30*, 99-120. doi: 10.1007/s11023-020-09517-8.

IEEE (2019). *Ethically aligned design. A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.* Retrieved from https://ieeexplore.ieee.org/document/8058187

Jobin, A., Ienca, M., & Vayena, E. (2019). Artificial Intelligence: the global landscape of ethics guidelines. *Nature Machine Intelligence, 1*, 389-399.

Knoef, M., Been, J., Alessie, R., Caminada, K., Goudswaard, K., & Kalwij, A. (2016). Measuring retirement savings adequacy: developing a multi-pillar approach in the Netherlands. *Journal of Pension Economics and Finance, 15*(1), 55-89. doi:10.1017/s1474747214000341

Lane, J. (2016). *A debt effect? How is unmanageable debt related to other problems in people's lives?* Retrieved from https://www.citizensadvice.org.uk/Global/CitizensAdvice/Debt%20and%20Money%20Publications/T he%20Debt%20Effect.pdf

Monserrate, S.G. (2022). "The Cloud Is Material: On the Environmental Impacts of Computation and Data Storage." *MIT Case Studies in Social and Ethical Responsibilities of Computing*, no. Winter 2022 (January). *https://doi.org/10.21428/2c646de5.031d4553*

Nibud (2017). *Keuzeproces bij financieel advies.* Retrieved from https://www.nibud.nl/onderzoeksrapporten/keuzeproces-bij-financieel-advies-2017/

Nibud (2019). *Geldzaken in de praktijk 2018-2019.* Retrieved from https://www.nibud.nl/wp-content/uploads/Nibud-Geldzaken-in-de-praktijk-2018-2019.pdf

Nibud (2022, July 15). Re: Nibud: groeiend aantal huishoudens zit financieel klem. Retrieved from https://www.nibud.nl/nieuws/groeiend-aantal-huishoudens-financieel-klem/

NOS (2022, September 6). Re: Hoogste inflatie ooit: 12 procent in augustus. Retrieved from https://nos.nl/artikel/2443478-hoogste-inflatie-ooit-12-procent-in-augustus

Ropohl, G. (1999). Philosophy of socio-technical systems. *Research in Philosophy and Technology, 4*, 186-194.

Social Science and Parliamentary Affairs Team, Houses of the Oireachtas (2010). Personal debt and consequences. Retrieved from: https://www.oireachtas.ie/parliament/media/housesoftheoireachtas/libraryresearch/spotlights/Personal_Debt_and_Consequences.pdf

Trist, E. L. (1981). *The evolution of socio-technical systems. A conceptual framework and an action research program.* Toronto: Ontario Ministry of Labour.

Van Dijk, W. (2016). *Grip op mentaal budget: een inaugurele rede.* Leiden, the Netherlands: Leiden University.

Vargo, S. L. & Lusch, R. F. (2016). Institutions and axioms: an extension and update of service-dominant logic. *Journal of the Academy of Marketing Science, 44*(1), 5-23.

Williams, J. (2018). *Stand out of our light: freedom and resistance in the attention economy.* Cambridge University Press.

Zhang, D., Mishra, S., Brynjolfsson, E., Etchemendy, J., Ganguli, D., Grosz, B., Lyons, T., Manyika, J., Niebles, J. C., & Sellitto, M. (2021). *The AI Index 2021 Annual Report.* Retrieved from https://arxiv.org/abs/2103.06312

OVERZICHT UITGAVEN
IN DE DESIGN PAPER SERIE