AI and Ethics
https://doi.org/10.1007/s43681-024-00468-9

ORIGINAL RESEARCH

Measuring adherence to AI ethics: a methodology for assessing adherence to ethical principles in the use case of AI-enabled credit scoring application

Maria Pokholkova[1] · **Auxane Boch**[1] · **Ellen Hohma**[1] · **Christoph Lütge**[1]

Received: 29 November 2023 / Accepted: 7 March 2024

Abstract

This article discusses the critical need to find solutions for ethically assessing artificial intelligence systems, underlining the importance of ethical principles in designing, developing, and employing these systems to enhance their acceptance in society. In particular, measuring AI applications' adherence to ethical principles is determined to be a major concern. This research proposes a methodology for measuring an application's adherence to acknowledged ethical principles. The proposed concept is grounded in existing research on quantification, specifically, Expert Workshop, which serves as a foundation of this study. The suggested method is tested on the use case of AI-enabled Credit Scoring applications using the ethical principle of transparency as an example. AI development, AI Ethics, finance, and regulation experts were invited to a workshop. The study's findings underscore the importance of ethical AI implementation and highlight benefits and limitations for measuring ethical adherence. A proposed methodology thus offers insights into a foundation for future AI ethics assessments within and outside the financial industry, promoting responsible AI practices and constructive dialogue.

Keywords Artificial intelligence ethics · Ethical assessment · Expert workshop methodology · AI-enabled credit-scoring · Transparency · AI ethics in finance

✉ Maria Pokholkova
maria.pkhva@gmail.com

[1] Institute for Ethics in Artificial Intelligence, School of Social Sciences and Technology, Technical University of Munich, Arcistrasse 21, 80333 Munich, Germany

Published online: 15 April 2024

1 Introduction

The misuse of decision-making artificial intelligence (AI) systems leads to unintended consequences stemming from the computational techniques and AI infrastructure employed in their development.[1] According to Ayling and Chapman [7], who focus on the epistemic concerns of technologies,[2] misuse stems from traditional data harms: non-intentional harms that result in individuals' problems associated with privacy violations,[3] discrimination,[4] and automatic consent.[5] Recently, a growing body of research addressed ethical compliance in the global AI landscape, encompassing Europe and beyond.[6] The authors also stress the need for practical tools that go beyond high-level ethical principles and focus on applying these principles in AI production and deployment, emphasizing the importance of addressing the "how" of applied ethics rather than just the "what."[7] Those efforts are translated, for example, into frameworks[8] in the public domain (recruitment, education, enforcement), which aim to systematically ensure that the high-level principles are operationalised.[9] However, Attard-Frost et al. [6] state that few AI ethical guidelines focus on fairness, accountability, and transparency within technical systems.[10] Jobin et al. [34] note, importantly, that the public's judgment tends to have a polarized view of AI algorithms, perceiving them as bad or good,at the same time, the ethical implications of AI technologies should be addressed on the level of design. That explains why assessing AI technology at every step of the AI lifecycle is important for tackling the problem of misuse.

Many frameworks, principles, protocols, and guidelines aim to evaluate the impact of AI technology and even provide standards for its quality in different domains. For instance, Value- Based Engineering (VBE), published by IEEE, prioritizes ethical considerations in designing AI systems.[11] Technical experts from the United States and China collaborate on AI technical standards globally.[12] However, governments struggle to cooperate on ethical AI standards, namely on the issues addressed, actors involved, and strategies used.[13] This fact postpones the establishment of global governance frameworks and results in a lack of interpretation of ethical AI rules and their operationalization for more detailed and concrete cases. This absence of a standardized approach to ethical AI has contributed to a global disparity in consumer trust in AI systems.[14] However, according to Omrani et al. [46], this trust can be enhanced by maximizing the technological features of AI systems. Hooks et al. [26] claim that the technological acceptance of various newly emerged AI-specific applications directly correlates to levels of trust in AI in general.[15] The research on trust in AI lacks in-depth examinations of specific AI cases and, specifically, the impact of the underlying trust factors on those cases. An analogy to this research problem can be illustrated with an example of research on evaluating the impact of Environment, Social, and Governance (ESG) reporting. Namely, research demonstrates that employing qualitative and quantitative methods effectively reveals the direct positive effect of ESG reporting on consumer trust in a company's brand, product, and service.[16] Consistent methods for quantifying compliance with the declared principles and values of AI systems' developers and deployers, including the mentioned high-level principles of ethics, become desirable and essential in establishing trust and promoting the adoption of AI systems.

Besides the challenges associated with integrating AI infrastructure within business organizations, which encompass both AI developed by businesses and AI employed within businesses, there are critical concerns related to ownership rights, cybersecurity, and data protection.[17] Moreover, those concerns should be tackled while achieving economic beneficence. According to the data collected from Western data sources, at least in Western literature, neither AI developers nor businesses using AI are obliged to follow jurisdictional rules or international AI principles.[18] Globally, the fragmented regulatory landscape, global variability, or variation of ethical considerations across countries, and the inherent complexity of AI raise the probability of emerging AI systems that risk harming humans. On national levels, a lack of a clear interpretation of political acts regarding the requirements for the ethics of AI increases the gap between public policy and the practice of using AI systems.[19] In these circumstances, assessing the level of trustworthiness in tools using AI systems seems, therefore, rather difficult due to

[1] Dwivedi et al. [14], Mills [40].

[2] Ayling and Chapman [7].

[3] Solove [56].

[4] Wachter et al. [67].

[5] Andreotta et al. [5].

[6] Amugongo et al. [3, 4], Fontes et al. [18], Corrigan [10].

[7] Ayling and Chapman [7], Morley et al. [43].

[8] Jobin et al. [34], Amugongo et al. [3, 4], Lütge et al. [36].

[9] Zhou et al. [69].

[10] Attard-Frost et al. [6].

[11] Spiekermann and Winkler [57].

[12] von Ingersleben-Seip (2023).

[13] Hagendorff [22], Hagendorff [23], Morley et al. [43], Hohma et al. [25].

[14] Omrani et al. [46].

[15] Hooks et al. [26].

[16] Tripopsakul and Puriwat [61], Koh et al. [35].

[17] Truby et al. [62].

[18] Truby et al. [62].

[19] Stix [58].

organizational and structural risks.[20] In terms of businesses, the lack of proactive strategy for integrating AI ethics into the corporate structure is explained by the "wait- and-see" policies arising from uncertainty.[21] It is why the current state of "ethical governance"[22] is underdeveloped. At the same time, the ability to effectively control the ethicality of AI systems constitutes a competitive advantage of such businesses, as it improves overall product quality and consumer trust.[23]

This paper seeks to contribute to the current research on AI ethics implementability by using the elicitations of experts, a technique already implemented to solve problems in political science, government, statistics, management science, and psychology. The elicitation of experts is asking a group of qualified individuals to express their opinions and judgments regarding uncertain events in terms of probabilities. According to the scientists who applied statistical methods to elicit expert knowledge, elicitation allows for incorporating subjective beliefs and opinions into probabilistic models. Despite the strong opinion that expert judgment cannot be quantified, and such a category as ethics cannot or should not be assessed quantitatively; statistical modeling in eliciting expert judgments proved effective in predicting complex physical phenomena.[24]

Due to the rapid entry of AI systems into the market, there is a pressing need for a self-consistent assessment method that would allow us to decompose the ethical characteristics of AI products and assess them. This problem can be addressed by involving qualified experts who can not only select features upon which to quantitatively evaluate the components of the overall composition but also consider the collective contribution of each feature to the overall assessment. Expert Workshop (EW), one of the approaches that employ statistical modeling, namely weighted sums, in expert judgment elicitation, fulfills these requirements. Publishing results from EW could attract and engage a broader range of experts in establishing quantitative assessment criteria for AI systems. Ultimately, this effort seeks to improve both the quality of AI systems and the performance of the financial companies utilizing them.

Therefore, in this paper, EW will be used to quantify the expert judgment of compliance with an ethical principle by an AI system that will be used for testing. This study aims to create a digital image of one of the characteristics of AI ethics used in the financial sector using data from a selected group of qualified experts. This paper illustrates the discussion using a case study involving an Expert Workshop where experts proposed a system of numerical criteria to assess the compliance of the AI Credit Scoring system with the principle of transparency.

Considering this, the hypothesis of this research can be formulated as follows: "Quantitative assessment of the constituent elements of AI ethics can be carried out based on a generalized expert opinion using statistical modeling (weighted sums) in expert judgment elicitation." Although our proposed metrics involve weighted sums of expert judgment, it is essential to recognize that this approach may have limitations compared to other potential methods. This paper will address these issues throughout the article, providing a comprehensive analysis of the strengths and weaknesses of the proposed methodology.

This paper will first provide a theoretical background defending the proposed methodology of the Expert Workshop for defining quantifiable measures of AI ethics principles. Subsequently, the outcomes of the proof of concept Expert Workshop will be presented, followed by a discussion on the usability and effectiveness of the proposed methodology.

2 Theoretical framework

2.1 Literature review on AI ethics measurement techniques

The analysis of AI ethics research can be presented in two categories: one that examines principles by comparing and classifying them without contextualizing them to a specific industry and another that analyzes techniques tailored to address the challenges related to AI ethics integration in particular sectors. An example of the first research category is the AI regulation strategy document, "Ethics Guidelines for Trustworthy AI," designed and published on behalf of the European Commission's AI High-Level Expert Group (AI HLEG).[25] The document is a non-obligatory framework within the European Union that implies the implementation of procedures into businesses that guarantee seven ethical principles: fairness, transparency, privacy, security, accountability, reliability, and safety.[26] Unlike other prominent high-level frameworks like the IEEE Global Initiative on Ethics[27] and the Montreal Declaration on Responsible AI,[28] AI HLEG focuses on the scope of Europe and promotes a regulatory approach that involves influencing regulatory bodies and policymakers when designing AI ethics implementation

[20] Koefer et al. [32].

[21] Framework Summary: Establishing a practical organizational framework for AI Accountability.

[22] Winfield and Jirotka [68].

[23] Morley et al. [43].

[24] Garthwaite et al. [20].

[25] HLEG [24].

[26] Radclyffe et al. [51].

[27] IEEE Global Initiative [30].

[28] Morandín-Ahuerma [41].

procedures. It comes to a second category of research encompassing an engineering approach of integrating ethics into the design of a concrete AI tool,[29] i.e., algorithmic decision-making. The research findings reveal that the ethical principles of fairness, transparency, and accountability are underrepresented in AI ethical business practices and, due to the disciplinary scope of ethics, are being replaced by speculative norms, i.e., corporate secrecy.[30] This suggests the need for more transparent methods to align AI ethics considerations with the business practices of organizations that deploy AI.

Understanding the difference between different quantification approaches that aim to assess the ethicality of AI systems is essential, as the optimal synergy between the most helpful assessment techniques is a must for strengthening the use of AI systems. For instance, literature distinguishes existing quantification frameworks that concern AI ethicality based on their efficacy, scope, focus, purpose, and manner in which they connect the cause and effect of the AI systems.[31] Among them are impact assessments, technology assessments, audits[32] tailored to an industry that involves AI ethics aspects, and design toolkits like value-centered design.[33] The impact assessments are also used for industry and business-specific purposes like achieving sustainable goals or ensuring stakeholder participation.[34] At the same time, there is no universally accepted rating of the most efficient or least efficient quantification methodologies for evaluating the integration of AI ethics principles into business practices. Given the evolving nature of the AI ethics field, adaptable and context-specific quantification methods play a valuable role. The method proposed in this paper aims to contribute to the ongoing dialogue and practical application of ethical principles in AI business contexts.

When defining the methods, AI research often focuses on qualitative evaluations of the adherence of AI systems to ethical principles or legal standards.[35] The assessment is also crucial for monitoring product quality, providing businesses with insights on enhancing product competitiveness in the market, and ensuring that consumers' rights are respected.[36] Among different options for a comprehensive evaluation of the ethicality of AI systems are ethical audits and assessments, frameworks, and guidelines developed by international,[37] national,[38] and industry-led initiatives[39] using interdisciplinary approaches. However, some practical guidelines or systems are designed to measure AI systems' ethical qualities, often proposed by ethical consultancies to ensure impartiality. For example, some publications evaluate AI systems' ethicality using a labeling approach introduced by AI Ethics Impact Group,[40] tailored to a concrete tool's specific context.

Another example is a TÜV SÜD AI Quality Framework[41] that proposes to measure risk for non- compliance to the legal framework for AI systems by calculating the severity of the AI ethical implication and the scope of the corresponding industry. These qualitative approaches are often regarded as practical decision-making tools with the potential to serve as monitoring tools for AI systems characteristics. They can be helpful to various stakeholders with different needs, including policymakers, regulators, and business owners.

The quantification allowed by those models is a competitive advantage that allows for control over the ethical quality of an AI system and, therefore, simplifies the organizations' harmonization of AI standards that will be outlined in the EU AI Act.[42] However, due to high complexity, the frameworks mentioned above can also be considered complex for understanding and implementation. Hence, the current state of AI ethical quantification frameworks must offer compelling evaluation examples. The quality of these assessments remains unclear, shielded behind non-disclosure agreements (NDAs) and corporate confidentiality. Finally, there needs to be evidence that those frameworks consider the importance of bringing various stakeholders to a consensus on the definition of ethical assessment.

A quantitative assessment of ethics in general, as well as the AI ethics and AI application's ethical adherence, is impossible due to the complexity and abstractness of these concepts as philosophical categories. At the same time, when using AI systems in practice, having a measurable level of trustworthiness as described in the HLEG [24] work and experts' opinions on such devices is helpful. Undoubtedly, the factors of AI systems that characterize trustworthy AI, including AI used in the financial sector, include ethics and integrity. At this stage of development and practical use of AI systems, it seems appropriate to decompose the general concept of ethics into components, a quantitative assessment of each of which can be carried out using the elicitation of expert knowledge. Merging databases of evaluations of elements, taking into

[29] Jobin et al. [34].

[30] Attard-Frost et al. [6].

[31] Ayling and Chapman [7].

[32] IEEE Standards (2019).

[33] Ayling and Chapman [7].

[34] Morrison-Saunders and Retief [44], Vakkuri et al. [66].

[35] Dolganova [12].

[36] Morley et al. [43].

[37] HLEG [24]>, IEEE Global Initiative [30].

[38] UK Parliament Committee [64], Executive Office of the President National Science and Technology Council [16].

[39] >TUV SÜD [63], Hallensleben et al. [21].

[40] Hallensleben et al. [21].

[41] TUV SÜD [63].

[42] Floridi [17].

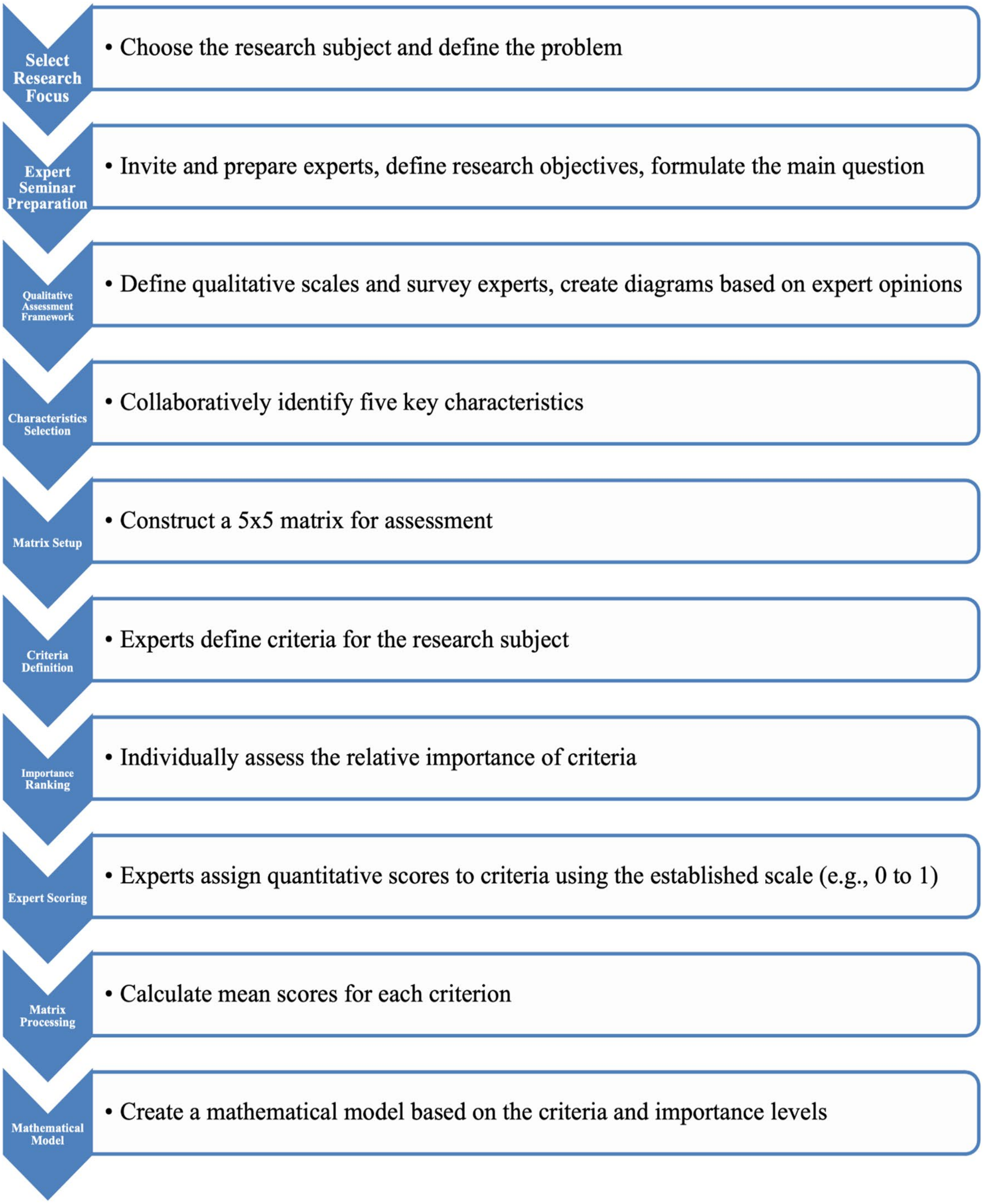

Select Research Focus
• Choose the research subject and define the problem
Expert Seminar Preparation
• Invite and prepare experts, define research objectives, formulate the main question
Qualitative Assessment Framework
• Define qualitative scales and survey experts, create diagrams based on expert opinions
Characteristics Selection
• Collaboratively identify five key characteristics
Matrix Setup
• Construct a 5x5 matrix for assessment
Criteria Definition
• Experts define criteria for the research subject
Importance Ranking
• Individually assess the relative importance of criteria
Expert Scoring
• Experts assign quantitative scores to criteria using the established scale (e.g., 0 to 1)
Matrix Processing
• Calculate mean scores for each criterion
Mathematical Model
• Create a mathematical model based on the criteria and importance levels

Fig. 1 The procedure of the expert workshop

account their importance, into a single database will allow us to obtain an "image" of the ethics of AI and consider its level when assessing its trustworthiness. An Expert Workshop (EW), a seminar-style approach that combines individual and group-based techniques to address complex ethical challenges or any problematic situations, leveraging the collective expertise of participants, seems a suitable methodology in this case.

2.2 Reusing the concept of expert workshop for quantification of adherence to ethical principles

The Expert Workshop (EW) is a method of seminar conduction that proposes a strategy to understand and quantify complex phenomena by involving 10–25 professionals who are specialized in corresponding phenomena.[43] This method was developed and proposed in a doctoral dissertation of Tolkacheva,[44] whose idea was to systematize a set of well-known practices and techniques for problem-oriented training for specialists in the field of engineering and technology. Two workshops were conducted using the EW methodology to assess students' adherence to Sustainable Development (SD) values[45] with Russian and international HEI (High Engineering Institutions) stakeholders, namely engineering students and educators. The experts were selected and invited to assess the engineering students' Sustainable Development mindset formation level. The characteristics chosen by both groups of experts were quantitative, relative, and applicable to any university and its engineering students. The authors also categorized the participants into three distinct groups: the level of the university's commitment to SD goals, SD mindset in the student community, and individuals' adherence to SD values. All characteristics represented a percentage or a share of the entity (i.e., "average % of study time within engineering courses devoted to SD issues").

Based on the expert assessment, it was found that the tested level of sustainability development (SD) mindset formation among engineering students in the investigated universities is low (73% of criteria). It suggests that 73% of the criteria used to evaluate the level of sustainability development (SD) mindset formation among engineering students have been met or fall into the "low" level category. At the same time, according to authors,[46] a comparison between the initial intuitive assessment and the subsequent quantitative assessment revealed that defining quantitative criteria and applying quantitative scales to the evaluation process led to a more comprehensive analysis; resulting in a more critical evaluation. Initially, the assessment indicated a much higher level of SD mindset formation, with 43.7–52.8% of responses suggesting a level above "low." In contrast, the later evaluation showed no indication of a level above "low."

The authors argue that a high-quality expert selection process for the EW is crucial for building a correct and comprehensive digital "image" of the problem. If a repeated EW is conducted, the probability of new evolving characteristics is close to zero. If the experts were selected correctly, it indicates that they are highly qualified, and the criteria they propose will likely align with or be similar to those suggested by other potential future experts. Nonetheless, with each successive EW conducted, the digital "image" of the phenomena becomes increasingly detailed. This process not only aids stakeholders in undergoing a conscious transformation but also fosters their inner motivation and understanding of the addressed transformation problem. A prerequisite of EW is the experts' competence, not based on their position or level of qualifications; but on their experience, direct involvement, and knowledge of the practical aspects of the problem.

The process for conducting expert research follows a structured sequence of steps (see Fig. 1). The preparation starts with selecting a seminar topic, subject, and research problem; followed by formulating requirements for experts and finally inviting chosen participants. Once experts are selected, they are provided with the seminar's goals.

The workshop begins with the qualitative phase, during which experts collaborate to adopt a definition, make assumptions, formulate the main question, and select a qualitative assessment scale. Individual surveys gather qualitative expert opinions, and expert teams nominate characteristics for quantitative assessment. In the qualitative phase, the moderation process plays a key role during the stage of characteristics nomination. The facilitator ensures that a consensus among the participants is reached. Moderated deliberation enhances the fairness of the characteristic selection and helps to conclude the discussion with the consent characteristic formulation among the participants. The moderator's task is to guide the discussion and maximize the consensus process within the group, facilitate precise phrasing of the characteristics, and help the group select the most argumentable and informative characteristics. At the same time, experts define their opinions and formulate improved characteristics by participating in the discussion. During the consensus process, voting can help if the deliberation and definition of opinions take too much time. Additionally, if the new characteristics are formulated by all participants based on the ones proposed by groups, they are written down and used during voting.

The most informative characteristics are selected during a participants' discussion, leading to the construction of a 5x5 matrix. This is followed by a quantitative phase of the workshop when criteria for the object's condition are established, and the comparative level of information content for each criterion is determined. Finally, experts provide quantitative assessments, which are mathematically processed to construct a model describing the subject of research based on selected criteria and their contributions. This process guides the transition from problem formulation to quantification and model creation.

2.3 Quantitative assessment and calculation process in expert workshop methodology

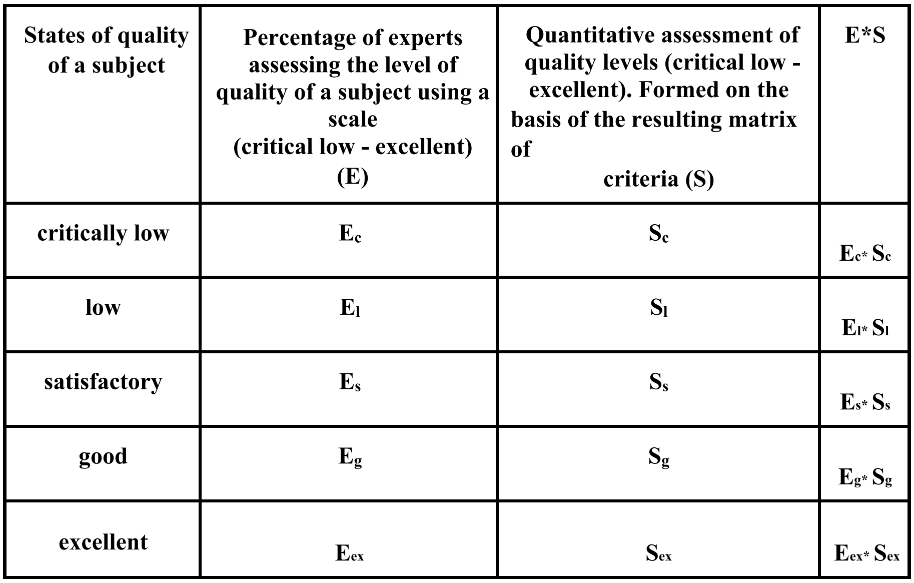
First, the "Aggregated Quantified Assessment" of the researched quality of a subject is calculated by multiplying each of the Status Quo values corresponding to its values

[43] Savinova [53].

[44] Tolkacheva [60].

[45] Pokholkov et al. [49].

[46] Pokholkov et al. [49].



States of quality of a subject	Percentage of experts assessing the level of quality of a subject using a scale (critical low - excellent) (E)	Quantitative assessment of quality levels (critical low - excellent). Formed on the basis of the resulting matrix of criteria (S)	E*S
critically low	E_c	S_c	E_c* S_c
low	E_l	S_l	E_l* S_l
satisfactory	E_s	S_s	E_s* S_s
good	E_g	S_g	E_g* S_g
excellent	E_{ex}	S_{ex}	E_{ex}* S_{ex}

Fig. 2 Assessment of subject quality and quantitative evaluation by experts

of Ratio of Importance. This assessment aims to provide a numerical value representing the level of subject quality. It uses relative values (KSQ1 to KSQ5) assigned to specific characteristics selected by experts to evaluate general quality (ranging from 0 to 1). Each characteristic's value is weighted by a ratio of importance (ɣ1 to ɣ5), where the sum of these ratios equals 1. It results in a generalized quantitative assessment of the subject's quality level, which adequately represents reality in the present moment according to the expert's perception. It is calculated as follows:

$$KSQ = (K1SQ * \gamma 1) + (K2SQ * \gamma 2) + (K3SQ * \gamma 3) + (K4SQ. * \gamma 4) + (K5SQ. * \gamma 5) \tag{1}$$

where KSQ1 … KSQ5 are calculated relative values of characteristics selected by experts to assess the current level of adherence of AI Credit Scoring to a principle of transparency (0–1);

$$\gamma 1 + \gamma 2 + \gamma 3 + \gamma 4 + \gamma 5 = 1 \tag{2}$$

where ɣ1 … ɣ5—the ratio of importance, or relative assessment of the specific weight of the selected criteria, within (0–1).

The Quantitative Assessment of the Levels of Qualitative States (QALQS) is calculated with K_i: a value of the criterion and ɣ$_i$: the specific weight of the *i*th criterion. Different qualitative states of quality are defined as: "Critically Low," "Low," "Satisfactory," "Good," and "Excellent." For each state, a formula is provided to calculate a numerical value based on the weighted sum of characteristic values (K_i) using specific weights (ɣ$_i$). Equations (3), (4), (5), (6), and (7) represent different thresholds of quality states, and the calculated values help to classify the subject quality into one of these qualitative states. It is calculated as follows:

$$S\ critically\ low = \sum (Kc.i * \gamma 1i) = (Kc.1 * \gamma 1) + (Kc.2 * \gamma 2) + (Kc.3 * \gamma 3) + (Kc.4 * \gamma 4) + (Kc.5 * \gamma t.5) \tag{3}$$

$$S\ low = \sum (Kl.i * \gamma 2i) = (Kl.1 * \gamma 1) + (Kl.2 * \gamma 2) + (Kl.3 * \gamma 3) + (Kl.4 * \gamma 4) + (Kl.5 * \gamma t.5) \tag{4}$$

$$S\ satisfactory = \sum (Ks.i * \gamma 3i) = (Ks.1 * \gamma 1) + (Ks.2 * \gamma 2) + (Ks.3 * \gamma 3) + (Ks.4 * \gamma 4) + (Ks.5 * \gamma t.5) \tag{5}$$

$$S\ good = \sum (Kg.i * \gamma 4i) = (Kg.1 * \gamma 1) + (Kg.2 * \gamma 2) + (Kg.3 * \gamma 3) + (Kg.4 * \gamma 4) + (Kg.5 * \gamma t.5) \tag{6}$$

$$S\ excellent = \sum (Kex.i * \gamma 5i) = (Kex.1 * \gamma 1) + (Kex.2 * \gamma 2) + (Kex.3 * \gamma 3) + (Kex.4 * \gamma 4) + (Kex.5 * \gamma t.5) \tag{7}$$

This calculation not only allows for a generalized quantified number for each of the states of quality but also accounts for the importance of each of the characteristics for the general quality of the subject.

The third calculated result, a Qualitative Expert Judgement (QEJ), is the result of the intuitive survey on the subject's current state. A scale from 0 to 1 to represent the qualitative judgments obtained from the survey. These judgments are expressed as shares, indicating the percentage of respondents who selected each qualitative category (e.g., critically low, low, satisfactory, good, excellent). The exemplary question for the survey can be: "What is, according to your opinion, the current quality state of the subject X?" For coherence with the numeric framework, the survey's answers are coded as E, and five E results, expressed as shares, match the quality states already used (critically low, low, satisfactory, good, excellent).

Finally, step four is a "Quantified Assessment of Average Collective Judgment of Experts," or QAACJE is done by multiplying the values of QEJ with the generalised scale or each of the respective states of quality QALQS (Fig. 2).

The final result is calculated in Eq. (8):

$$QAACJE = (Ec*Sc) + (El*Sl) + (Es*Ss) + (Eg*Sg) + (Eex*Sex) \tag{8}$$

This multiplication results in the formation of a new scale that unites both qualitative and quantitative perceptions of experts, and the summation of those values gives a number that summarizes the expert's assessment of the subject's quality. It is important to clarify that QAACJE is a valuable component of the Expert Workshop methodology, as it is a technique that enhances its flexibility and effectiveness in transforming opinions into quantifiable data. However, the method of weighted sums does not eliminate the subjectivity of expert opinions. The strengths and weaknesses of the EW procedure and its metrics will be discussed in the next chapter.

2.4 Exploring the nuances of expert workshop (EW)

According to Morgan [42], poorly done expert elicitation, when used for applied decision analysis, can discredit the whole approach and lead to useless or deceptive results. Moreover, the elicitation procedure should account for inherent biases and minimize them within the process and in the results. Therefore, some principles and interpretations of the methods and techniques and the concept might be necessary for the reproducibility of the EW methodology. Finally, the methodology's weaknesses should be minimized by raising awareness about its shortcomings and explaining the measures of control over the method.

2.4.1 The principles of EW preparation and conduction

One of the most crucial steps in organizing an EW is the selection of experts. Experts are professionals of a specific field with accumulated knowledge and expertise, complemented by their deep understanding of the subject's constraints and advantages. Therefore, selecting appropriate experts is based on three principles: qualifications in a particular field of research, a high level of engagement, and professional interest in finding a solution to the problematic situation addressed in the workshop. Organizers of the EW ensure that at least two of the principles should be satisfied when selecting experts:

The principle of **Relevance** is implemented through the study of publications, information about conferences, seminars, and other events that allow the identification of a pool of qualified researchers on the relevant issue who may subsequently be invited to participate in the EW. For example, the invited experts should have at least one publication on the subject of investigation in the last three years or a minimum of two years of work experience in the context of the subject.

The principle of **Engagement** is realized by inviting experienced individuals who have knowledge about the phenomenon studied in the EW from their professional activities or everyday lives. Often, the expert opinions of such individuals are no less valuable than those of qualified expert researchers.

The principle of **Motivation**: individuals show motivation to resolve the problem of the phenomenon under study. This is particularly important due to the necessity of collectively finding ways to resolve the researched problem during the EW seminar.

Regarding the principles of EW conduction, facilitators play a crucial role in maintaining neutrality towards the various perspectives of experts. This vital principle of EW conduction is to ensure that experts feel comfortable expressing their opinions without feeling pressured to adopt a dominant viewpoint and to create conditions for experts to express their ideas.

2.4.2 Comparison with other expert judgment elicitation methods

Several categories of methods can be distinguished among numerous publications on expert elicitation methods, or methods of gathering the insights and opinions of knowledgeable individuals in a particular field regarding uncertainty. Many are expert elicitation methods explicitly tailored to the public sector,[47] environmental science and risk assessment,[48] policy analysis,[49] etc. Even though some techniques declare their ability to assess phenomena,[50] it is unclear whether there is an effective comparable method to the EW that specializes in assessing the phenomena's state.

In the context of Expert Judgment Elicitation (EJE) taxonomy, a categorization system that organizes various methods and approaches used in expert judgment elicitation, EW could be attributed to quantitative and qualitative methods that use fluent and numerate methods.[51] Fluent methods involve gathering qualitative or descriptive information from experts, which aim to capture the experts' subjective

[47] Butler et al. [9].

[48] Usher and Strachan [65].

[49] Morgan [42].

[50] Hsu [27].

[51] Szwed [59].

insights, opinions, or experiences without quantifying them into numerical values. Numerate methods aim to provide more precise assessments and can include probability estimations. The concept of the EJE provides a foundation that describes the EW method, which combines direct and indirect elicitation and individual and consensus aggregation.[52] Therefore, according to the EJE taxonomy, EW is a mixed-method group elicitation approach that combines qualitative expert judgment methods with quantitative methods, mathematical methods, or a weighted factor method.[53]

EW could be compared to the Delphi method, a consensus-building technique that uses questionnaires to collect participant data.[54] However, the Delphi method often uses the opinion of geographically dispersed experts.[55] Therefore, it builds on electronic and anonymous communication, which does not leave room for clarification when interpreting the results. In contrast, EW allows for collaborative face-to-face interactions among experts, facilitating the development of agreed-upon judgments and the selection of informative numerical criteria. Delphi consists of 3–4 iteration rounds in which experts must give their statements and then reassess them to reach a consensus at the end of the process. Compared to EW, Delphi's method presents design vulnerabilities. Delphi's method has no requirement of being present and engaged, and the method includes the obligation to grant participants a large block of time (i.e., 2 weeks).[56]

Another method, the Analytic Hierarchy Process (AHP), is a mixed-method approach that uses pairwise comparisons to derive weighted comparisons. This method can be better compared to EW, as both EW and AHP use literate and numeric metrics. AHP is a practical decision-making method that divides complex problems into hierarchical structures, allowing for comparing elements and calculating weights based on expert judgments and the relationships between factors.[57] Specifically, the similarity with AHP is noted as both AHP and EW, apart from freeform methods, such as brainstorming, use scaling methods with discrete and continuous ratings (i.e., 0–1).

According to the literature review, EW is a unique method for gathering expert judgments and selecting informative numerical criteria. It is distinct from other established methods like AHP in its focus on assessing the state of a problem or phenomenon rather than choosing among alternatives. Thus far, evidence suggests that expert elicitation methods primarily address selecting options from multiple alternatives. While EW can be formally compared to AHP as they both employ quantitative metrics, the applicability of weighted sums metrics for expert judgment elicitation requires further research.

2.4.3 Accuracy and metrics (weighted sum method)

Human brains struggle to process large amounts of data or perform intricate statistical computations,[58] therefore, there is no technical possibility to validate the accuracy of the results obtained while eliciting expert judgment. Uncertainty must be accepted when judging the probability of events and the inherent cognitive biases. The EW demonstrates the aspect of statistical representation of the expert knowledge. It states that the accuracy and statistical significance of the results obtained depend on the level of competence and the number of experts involved. Further efforts to expand the amount of experts and raise the competence level of the experts involved would increase the accuracy of the obtained group result. However, there should be a limit of approximately 25 people to keep the discussion engaging and manageable.

The structure of the matrix approach exemplifies the necessity for a participation number limitation. The accuracy of the digital portrait of the investigated phenomenon depends on the number of selected characteristics for its description and the numerical criteria for evaluation. More characteristics mean a more detailed and accurate digital portrait of the phenomena; this also applies to the range of qualitative assessment scales. The Likert Scale is the qualitative scale used for the Expert Workshop; it proposes a scale of five states of the phenomena, providing standardization and accuracy across assessments.[59] At the same time, the greater the number of features and the wider the range of qualitative assessments, the more work the experts need to do during the seminar. The matrix approach (used to obtain a digital image of the investigated phenomena when using a 10 × 10 matrix) may allow for the exclusion of maximum and minimum values during statistical processing but significantly increases the duration of the seminar. A 5 × 5 matrix allows for conducting a workshop with 15–20 participants in 3.5–4.0 h. However, a 10 × 10 matrix would take at least 8 h. According to previous tests of the methodology and the experts' feedback, an acceptable level of accuracy of the obtained digital image of the subject under investigation is achieved with a 5 × 5 matrix.[60]

Apart from the quality of experts, the selection of metrics directly impacts the accuracy and reliability of the results obtained through expert judgment elicitation and helps to reduce bias. Multi-criteria problems are fundamentally more

[52] Szwed [59].

[53] Szwed [59], Satybaldiyeva et al. [52].

[54] Hsu [27].

[55] Adams [1].

[56] Delbecq et al. [13], Hsu and Sandford [28].

[57] Saaty [54].

[58] Morgan [42].

[59] Pasman and Rogers [48].

[60] Pokholkov et al. [49].

complex than single-criteria problems and require unique methods to find their solution.[61] The Expert Workshop aims to conduct a multicriteria assessment, for which the weighted sum method (WSM) is convenient. The Weighted Sum Method (WSM) is an approach used in decision-making that aggregates multiple criteria into a single composite criterion; typically represented as a weighted sum of the individual criteria.[62] Namely, calculations that are applicable for decision-making tasks in various scenarios; such as the selection of the best option or multiple best options, ordering all options by preference, and assessing the characteristics. However, solving multi-criteria problems, such as those in multi-criteria decision analysis (MCDA), requires significant effort to gather and process decision-makers' preferences. This can be resource-intensive and time-consuming. Moreover, since MCDA relies on subjective decision-maker preferences, there are no objective solutions for comparison; posing challenges in evaluating results against benchmarks.

The simplicity of the 0–1 continuous scale used in the Expert Workshop quantification phase is beneficial because it allows for an intuitive assessment process. Despite the complexity of the studied problem, which involves ethical considerations; this approach streamlines the evaluation to make it easier for experts to provide their insights. In ethical quality assessment, where multiple factors and perspectives are at play, a simple system helps condense the core ideas of experts' qualitative judgments into quantifiable measures to facilitate a clearer understanding of the overall ethical landscape. However, the WSM's dependency on expert judgment, namely on their subjective opinion, can introduce potential biases or errors in the subject's assessment.[63] At the same time, while being biased, the method allows for the quantification of subjectivity. This accounts for the direction in which the subjectivity of a specific group of experts is directed.

Understanding subjectivity could determine the stakeholders' priorities in the topic of development.

2.4.4 Group subjectivity in expert judgment

In an Expert Workshop (EW) scenario, the subjective probability expressed by an expert reflects their personal belief; which is influenced by both formal evidence and informal knowledge or experiences. Despite being biased, subjective probability distributions (SPD) are more effective than other statistical methods when eliciting uncertain expert knowledge.[64] In a subjectivist or Bayesian perspective, individuals assess the likelihood of uncertain events or quantities based on their subjective judgments about the present or future state of the world and the underlying governing processes.[65] Regarding group elicitation, the subjectivity extends to the group dynamics in EW.

As a result, collective judgment is influenced by groupthink—when group members feel peer pressure to conform with a dominant opinion. This is reinforced by cultural stereotypes, such as the ultimate importance of group members' opinions who outweigh others regarding their social status (age, gender, authority, and professional achievements) during the workshop. Another consideration during the process includes the moderator's awareness of group dynamics, ensuring that each expert holds equal weight during the discussion.

At the same time, consensus building, an essential activity of Expert Workshop (EW) achieved due to its in-person setup, aims to reach an agreement through open dialogue, negotiation, and compromise among participants. Various methods and techniques employed in EW prioritize avoiding groupthink and achieving consensus. Revising the group work results during the consequent critical analysis of the whole group of experts contributes to a more precise understanding of elicitation concepts, notions, or elicitation questions. Notably, the multi-criteria decision analysis process helps to ensure that all relevant factors are named, formulated, and considered. The success of Expert Workshops hinges on balancing individual subjectivity, group dynamics, and effective consensus-building strategies to ensure accurate and reliable collective judgments.

2.4.5 Challenges associated with engaging experts

Contacting experts via email using organizers' professional and personal networks, conducting "cold calls" using online databases and social media, and confirming the interest of experts can be challenging. It requires extensive effort to reach out to suitable experts and ensure their participation, as usually, it is difficult for a sufficient number of experts to be physically present on a specific date, time, and place. Additionally, hosting workshops in a fixed location may limit the involvement to only those who can physically attend. However, this limitation can be mitigated by coordinating workshops with other events that draw experts from diverse locations. This challenge is also mitigated when seeking a localized perspective on the investigation.

At the same time, the challenge remains to find experts that would fit the investigated subject. The methodology procedure is explained in the invitation, which allows experts to decide on participation. However, even with the detailed explanations in the invitation, some experts may not fully

[61] Podinovski and Potapov [50].

[62] Podinovski and Potapov [50].

[63] Podinovski and Potapov [50], Garthwaite et al. [20].

[64] Lenthe [33].

[65] Garthwaite et al. [20].

agree with the contextual definition provided by the organizers, which may lead to disagreements during the workshop. For instance, according to the past use cases of EW,[66] it is likely that at least one expert in the group refuses to accept the contextual definition proposed by an organizer. In this case, the moderator asks the expert to provide their definition. So far, there has not been a case where the contextual definition of the investigated subject of the EW has been formulated promptly by an expert.

2.4.6 Stakeholder dynamics

Another challenge associated with stakeholder dynamics occurs when the stakeholders play different roles in the process (i.e., seller and buyer, user and creator). Achieving consensus during the workshop can be challenging due to the risk of stakeholders leading polarized discussions. Moreover, briefing and moderation must mobilize the participants to ensure that experts remain engaged, thoughtful, and productive. However, it can be difficult for a moderator to manage the group due to the need to accommodate participants' diverse backgrounds and preferences. Moreover, a moderator should be experienced and knowledgeable enough to coordinate smoothly and promptly during all stages of the EW; reacting appropriately to experts' questions, comments, and objections. During the group work stage, where experts collaborate to formulate characteristics, the moderator must monitor multiple groups simultaneously; to ensure that discussions progress in the right direction. For example, developing quantifiable characteristics can challenge the participants. If a group encounters difficulties, the moderator should react proactively and provide guidance by naming correctly formulated characteristics. Several instructed moderators instead of one could simplify the moderator's task in cases when an EW hosts a large number of experts or organizers who lack experience in workshop conduction.

Since the EW method relies more on qualitative assessments and expert consensus than complex mathematical models, it exclusively addresses real-life problems and practical challenges with complex, multifaceted contextual nuances, like the ethics of AI. Considering this article aims to find a suitable methodology for hands-on, functional, measurable characteristics of AI systems' ethics, the preferred method would be interdisciplinary to provide diverse perspectives. In this context, the use case of credit scoring is selected as the AI system for testing as it holds significant relevance in the financial sector. Credit scoring systems are widely used in lending and financial decision-making processes; impacting individuals and businesses. Given the potential implications of biased or unethical AI algorithms in this domain, the choice of AI CS as the use case allows for a focused examination of a real-world, high-stakes application of AI ethics.

3 Procedure of the proof-of-concept workshop

3.1 The use case of credit scoring

The integration of Artificial Financial Intelligence (AFI) into the operations of modern fintech companies and traditional financial organizations is a rapidly growing trend. Artificial Financial Intelligence (AFI) refers to AI techniques and technologies that automate financial processes while complementing existing human financial expertise.[67] Modern fintech companies and traditional financial organizations that aim to upscale their financial operations undergo a snowballing process of AFI integration[68] in their business models. One example of AFI is AI- enabled credit scoring (AI CS). This AI type of program automates and replicates aspects of human financial expertise through a combination of machine learning (ML),[69] a subcategory of AI algorithms and other AI techniques. From an economic and technical perspective, Credit Scoring (CS) using AI has brought forth a spectrum of applications, such as Machine Learning

(ML) and sophisticated Deep Learning (DL) methodologies like Neural Networks, known for their proficiency in deciphering complex data relationships.[70] The selection of a particular model hinges upon several critical factors, including the scale and quality of available data, the intricacy of credit decisions, and the specific requirements of lending institutions.[71] AI CS can use financial and non-financial data sources, including social media activity, textual data, and online behavioral patterns. The abovementioned AI models excel compared to humans in assessing the probability of loan repayment.[72] Lastly, AI CS can autonomously decide whether to grant individuals or entities a loan or other financial services.

The advantages and disadvantages of AI CS, including its potential for improved credit risk assessment, enhanced financial policies, and concerns about unconscious bias, have been extensively discussed in the recent literature on

[66] Pokholkov et al. [49].

[67] Pokholkov et al. [49]

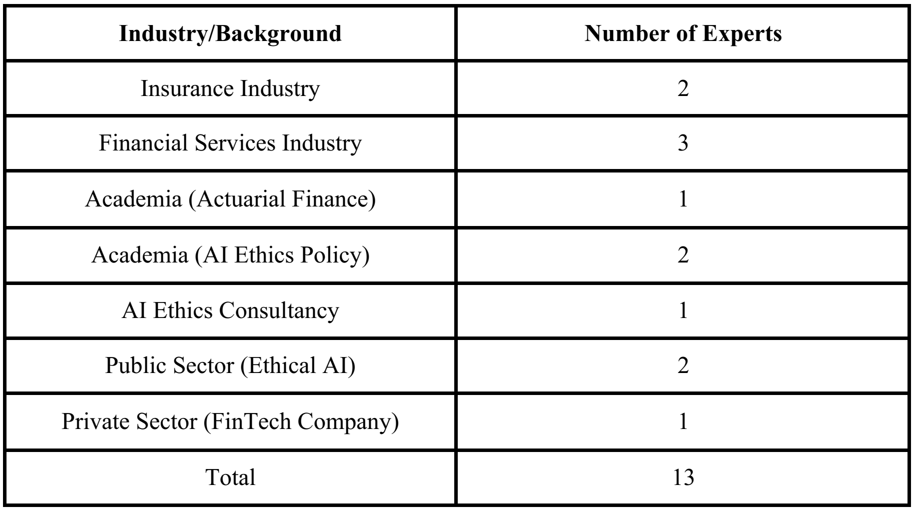
[68] Solanki [55].

[69] Kumar et al. [37].

[70] Huang et al. [29].

[71] Eddy and Bakar [15].

[72] Ben-David and Frank [8].



Industry/Background	Number of Experts
Insurance Industry	2
Financial Services Industry	3
Academia (Actuarial Finance)	1
Academia (AI Ethics Policy)	2
AI Ethics Consultancy	1
Public Sector (Ethical AI)	2
Private Sector (FinTech Company)	1
Total	13

Fig. 3 Composition of Expert Workshop Participants

trustworthy AI.[73] From an academic perspective, AI CS is described as a powerful tool for financial inclusion, providing affordable service to vulnerable members of society through algorithmic decision-making that mimics intelligent human behavior.[74] At the same time, while AI CS demonstrates cost-efficiency compared to traditional human creditworthiness assessments and the potential for financial inclusion for borrowers, it also raises significant ethical concerns that need careful consideration[75]: potentially discriminative AI decisions, which are based on biased algorithms; lack of interpretation on how a decision was made. In the literature, some methods are described that allow for evaluating the AI CS effectiveness, predictability, and justification mechanisms underlying the tool's decision.[76] Specifically, in the literature examining the economic angle of AI CS when analyzing artificial intelligence, more attention is paid to economic aspects, such as the accuracy of AI CS risk evaluations.[77] However, in several specific cases, using AI in the financial sector raises serious ethical issues that require careful consideration, such as compliance with fairness, accountability, and transparency.[78] Recently, many researchers have emphasized prioritizing the principles of fairness and transparency over other moral principles.[79] This fact suggests that the change of question from "what" to "how" in the applicability of AI ethics is especially relevant for the specific AI industry in finance.

Regarding the impact assessment of AI technologies in finance, solutions have been mentioned in academic and industry dimensions. Organizations and fintech companies across various industries, from technical equipment providers offering AI-powered platforms to financial consultancies involved in the entire AI CS production cycle, have devised their practices, guidelines, and metrics for adhering to ethical AI principles. These frameworks have been developed, for instance, by IBM (AI Fairness 360),[80] Ernst and Young (Trusted AI Framework),[81] and JP Morgan Chase (Explainable AI Centre of Excellence).[82] These initiatives have not been developed explicitly in response to the AI HLEG's recommendations. However, they share a broader objective of promoting responsible and ethical AI practices. Frameworks are committed to addressing principles such as fairness, transparency, and accountability in the context of companies' services and product characteristics. For instance, the IBM framework is declared to detect bias in the machine learning models used to train AFI and AI CS and remove them, as the reliance on biased algorithms potentially leads to unfair or discriminatory decisions. Such bias can manifest in various forms, including prioritizing certain user groups based on ethnicity, gender, or income. Even though AI CS application scenarios have employed a quantitative assessment to measure the fairness of decisions made by AI CS to account for the extent to which its decisions are fair regarding different demographic groups, there are no signs that

[73] Curto et al. [11].

[74] Ozili [47].

[75] Maree et al. [38].

[76] Kumar et al. [37].

[77] Ghodselahi and Amirmadhi [19].

[78] Max et al. [39], Ahmed [2], Nowakowski and Waliszewski [45], Kozodoi et al. [36].

[79] Kozodoi et al. [36].

[80] IBM Developer Staff, "AI Fairness 360" https://www.ibm.com/opensource/open/projects/ai-fairness-360/ (2018).

[81] Ernst and Young Staff, "Responsible AI", https://www.ey.com/en_ch/ai/responsible-ai, n.d.

[82] JP Morgan Chase, Explainable AI Centre of Excellence, https://www.jpmorgan.com/technology/artificial-intelligence/initiatives/explainable-ai-center-of-excellence, 2023.

those metrics are widely applicable.[83] For example, limited evidence suggests that these policy frameworks comprehensively address organizational challenges in ensuring AFI's compliance with the ethical criteria established by the EU AI HLEG [24]. At the same time, considering the speed of AFI integration, there is a pressing need for an understandable, unified methodology that includes metrics that can evaluate adherence to the most challenging ethical concerns surrounding AFI. The need for transparency in AFI was also articulated by a preliminary survey conducted on the preparatory stage of an Expert Workshop: transparency was selected to be the second most important ethical principle for AFI after security. Therefore, the importance of transparency is underscored by the choice of this principle for a proof-of- concept workshop.

In summary, AI-enabled credit scoring (AI CS) is rapidly reshaping the financial industry, offering improved credit risk assessment while raising critical ethical concerns. Given the paper's objective to demonstrate the feasibility of the Expert Workshop (EW) method in quantifying adherence to ethical principles, the selection of AI CS for the proof of concept is adequate: AI CS provides an ideal example of a combination of financial and technical aspects. Also, considering the necessity to decompose the general concept of ethics on its underlying principles and test them separately, it was decided to methodologically test the principle of transparency based on its definition of the AI HLEG framework. Furthermore, the choice of test principle was informed by the results of a preliminary survey conducted remotely among the invited experts before the EW venue. This survey demonstrated the overarching importance of the transparency principle in the AFI. Also, the interdisciplinarity of the AI CS use case matches the specificity of the Expert Workshop methodology that fosters an interdisciplinary approach for effectiveness in assessing ethical concerns in a specific domain.

3.2 Proof of concept workshop

3.2.1 Participants

The pre-selection of experts was done via search through social network engines, such as LinkedIn, the research of thematic forums, academic search platforms, and databases focused on AI in finance. The requirements for candidates were their professional employment in the AI financial sector. AI developers, ethicists, financial professionals, regulators, consumer advocates, and business and academia representatives who possess specific knowledge about the decision-making process of AI in finance were considered suitable candidates for the EW in AI CS. Additionally, experts with knowledge of decision-making in traditional finance and individuals who work with organizational risks for ethical AI implementation were welcomed. Of 55 invited experts, thirteen confirmed their participation in the Workshop: five women and eight men (Fig. 3).

3.2.2 Preliminary survey on the ethical principles of AI systems in finance

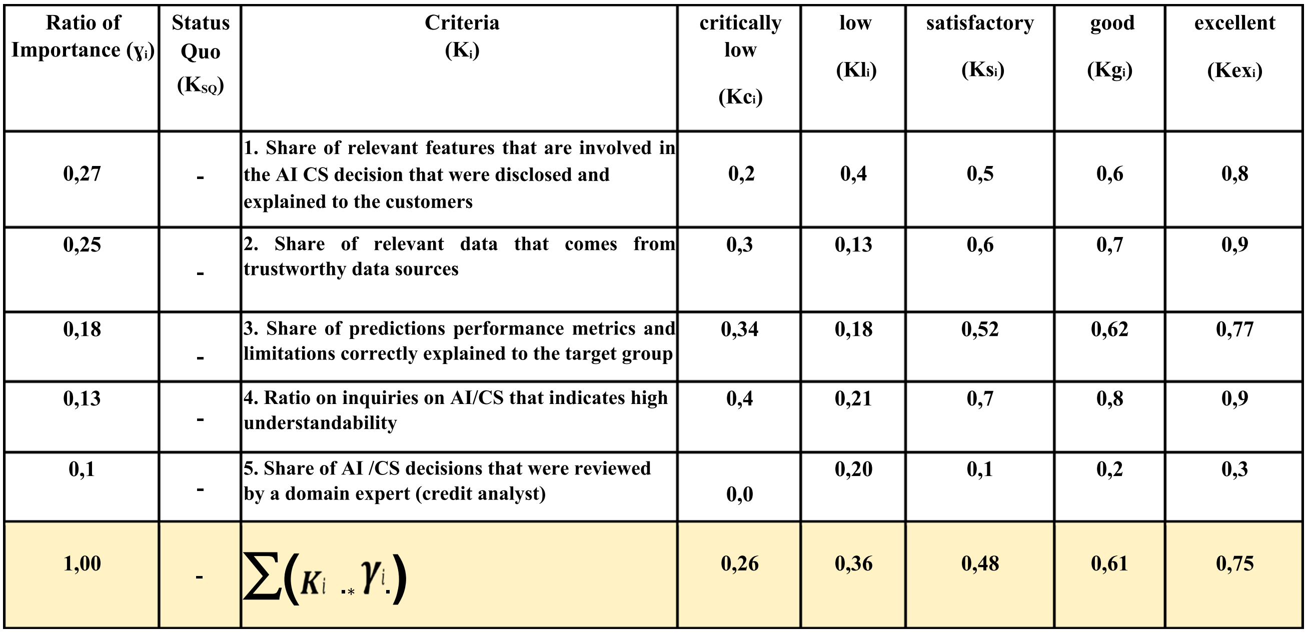
The choice of testing the adherence of the AI CS system to only one most important ethical principle, rather than a set of principles, is explained by the structure of the Expert Workshop methodology that allows for the decomposition of complex phenomena into more minor elements. In the seminar context, assessing the ethical quality of AI systems in finance proves challenging due to the multifaceted nature of ethical principles involved in this domain. This implies that by focusing on one principle, one can thoroughly analyze and evaluate its application in the context of AI systems in finance. The purpose of the preliminary survey was explained by the necessity to define the most pressing ethical problem about the AI systems employed in the financial industry and test it using expert knowledge.

Selected candidates for the workshop were invited via email to vote on the most important ethical principle in AI in finance based on the five principles (transparency, fairness, privacy, security, and accountability). The survey question "Which two characteristics are the most important ones in AI systems when applied in the finance industry?" was answered by six people, and security scored four, the most considerable number of votes. Three votes were equally scored by fairness and transparency, whereas privacy scored two votes and accountability scored one vote. Considering the totality of the factors, such as the limited reachability to the experts with the necessary knowledge and the need to align the workshop research focus with the contextual knowledge of confirmed participants, as well as the prevailing literature discussions on AI ethics in AFI, transparency was chosen as the principle to be tested.

3.2.3 Expert workshop

As a first step, participants were presented with background research on the problems of the ethical aspects of AI Credit Scoring and the methodology of the Expert Workshop. Additionally, experts were presented with the preliminary survey results on the most important ethical principles and proposed using the contextual understanding of ethicality during the workshop. Also, they were asked to consent to the predefined definitions and conditions for the common usage in the context of the EW. Namely, participants had to agree with the validity of the following statements: "An AI CS is considered ethical if it has adhered to the principles of

[83] Jammalamadaka and Itapu [31].



Ratio of Importance (γ_i)	Status Quo (K_{SQ})	Criteria (K_i)	critically low (Kc_i)	low (Kl_i)	satisfactory (Ks_i)	good (Kg_i)	excellent (Kex_i)
0,27	-	1. Share of relevant features that are involved in the AI CS decision that were disclosed and explained to the customers	0,2	0,4	0,5	0,6	0,8
0,25	-	2. Share of relevant data that comes from trustworthy data sources	0,3	0,13	0,6	0,7	0,9
0,18	-	3. Share of predictions performance metrics and limitations correctly explained to the target group	0,34	0,18	0,52	0,62	0,77
0,13	-	4. Ratio on inquiries on AI/CS that indicates high understandability	0,4	0,21	0,7	0,8	0,9
0,1	-	5. Share of AI /CS decisions that were reviewed by a domain expert (credit analyst)	0,0	0,20	0,1	0,2	0,3
1,00	-	$\sum(K_i \, .* \, \gamma_i.)$	0,26	0,36	0,48	0,61	0,75

Fig. 4 Matrix of criteria for assessing the level of transparency of AI Credit Scoring (on a scale of 0–1)

ethical AI, particularly transparency." Also, "The ethicality of the AI Credit Scoring tool can be qualitatively assessed by measuring its adherence to the qualitative characteristics inherent to AI CS products."

The first step in the Ethical Workshop (EW) also involved simplifying the concept of AI ethicality by equating it with transparency. This simplification was deliberately designed to enhance contextual comprehension of ethicality, and the reason for this simplification is rooted in the intricate nature of ethical considerations, which often originate from the realm of philosophy and are prone to individual and, consequently, biased interpretations. The introduction of these contextual definitions and conditions marked a pivotal initial stage in the methodology, aimed at testing both the hypothesis's validity and the validity of the eventual results. Establishing these conditions is critical as it streamlines the intricate landscape of ethical considerations and fosters consensus among the workshop participants.

After agreeing on using contextual definitions in terms of the EW, experts were invited to individually share their opinion on the current state of AI Credit Scoring adherence to transparency using an online multiple-choice survey. Namely, they were given five types of answers: excellent, good, satisfactory, low, and critically low state to characterize the subject. The results were displayed in the form of a diagram for the attention of all experts specifically. The majority of experts considered the AI CS transparency to be low.

As a next step, the Expert Workshop (EW) participants were divided into teams, with two groups consisting of four experts each and one group comprising five experts. Each group was tasked to name five measurable characteristics that would allow for the qualitative assessment of the AI Credit Scoring tool's transparency. For that, participants received handouts[84] providing a context for ideation of the characteristics of AI CS. Namely, participants were given an example of an abstract AI CS tool developed and tested in the EU and based on Artificial Neural Networks (ANN) models. Also, it was mentioned that the producer company claims that their tool expands access to capital and financial services for marginalized communities and uses financial and non-specified alternative data for decision-making when the client consents to disclose its data, as required to comply with GDPR.

Each group of experts was proposed to name such characteristics or features that should be evaluated on a scale from 0 to 1. After 15 characteristics were named in total (see Appendix B)[85] participants were invited to participate in a quorum discussion with other groups to select five of the fifteen most relevant characteristics and formulate them to be scalable from 0 to 1. As a part of the process, group representatives had to defend their formulation of characteristics and consider criticism of other groups. This stage of the EW took the most significant share of the total duration of the EW. Specifically, all 13 participants were challenged to agree on the formulation of scalable characteristics, as

[84] Appendix A: Handout for Participants.

[85] Appendix B: 15 Characteristics.

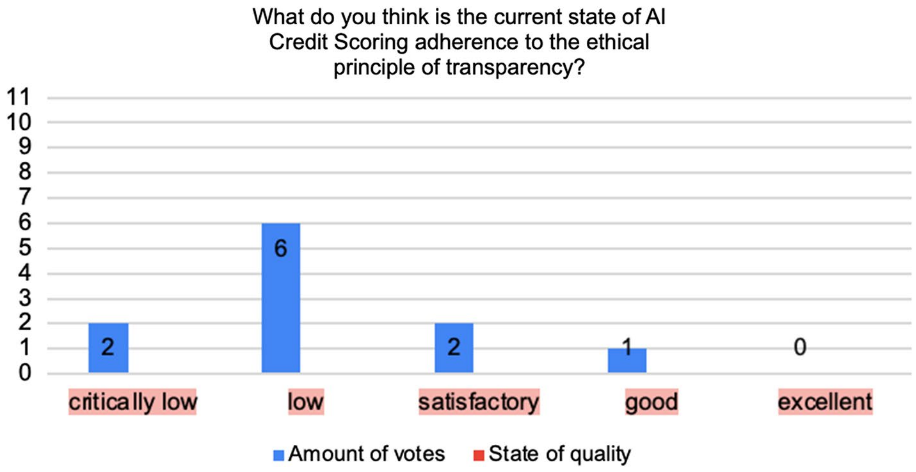
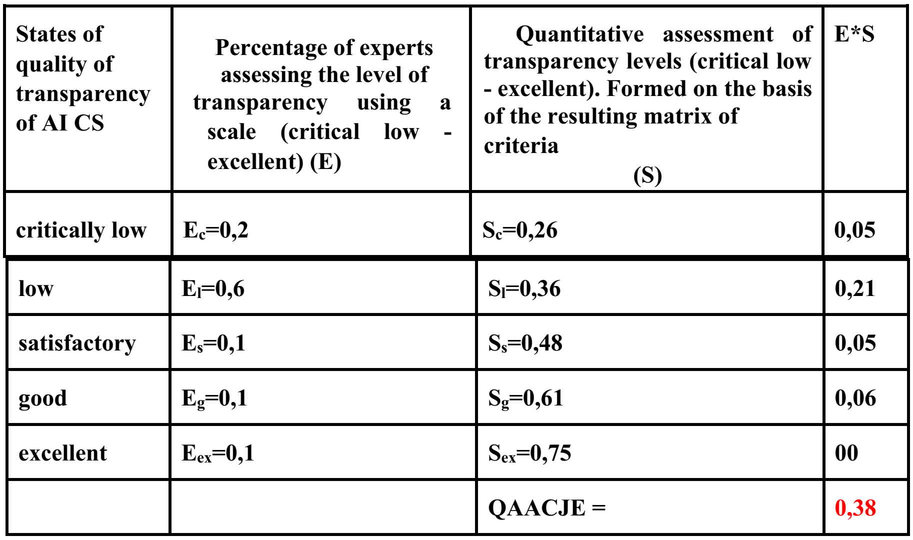


Fig. 5 Intuitive survey results



States of quality of transparency of AI CS	Percentage of experts assessing the level of transparency using a scale (critical low - excellent) (E)	Quantitative assessment of transparency levels (critical low - excellent). Formed on the basis of the resulting matrix of criteria (S)	E*S
critically low	E_c=0,2	S_c=0,26	0,05
low	E_l=0,6	S_l=0,36	0,21
satisfactory	E_s=0,1	S_s=0,48	0,05
good	E_g=0,1	S_g=0,61	0,06
excellent	E_{ex}=0,1	S_{ex}=0,75	00
		QAACJE =	0,38

Fig. 6 Assessment of AI CS transparency and quantitative evaluation by experts

their perspectives on what constitutes transparency factors for AI CS did not align.

With that selection, the five best characteristics were inserted into the matrix table in Microsoft Excel,[86] and the table was shared with all participants individually. Experts were asked to work individually, filling the matrix using their expert knowledge. Namely, participants were tasked with assigning values from a scale of 0 to 1 to evaluate five characteristics, a ratio indicating the importance of each characteristic, and a scale representing the status quo of the AI CS in transparency. The personal Excel table was shared with each participant via a weblink, allowing them to input the numbers they consider adequate and consistent with their expert knowledge. The values proposed by each expert were processed with thirteen Microsoft Excel Lists and connected to one standard matrix programmed to calculate the arithmetic mean of each criterion. Due to technical and organizational challenges, only ten participants could complete the matrix. As a result, the quantitative results for the group are based on the assessments provided by these ten experts, as defined in Fig. 4.

Industry/background	Number of experts
Insurance industry	1
Financial services industry	2
Academia (actuarial finance)	1
Academia (AI ethics policy)	2
AI ethics consultancy	1
Public sector (ethical AI)	2
Private sector (FinTech Company)	1
Total	10

[86] Appendix C: Microsoft Excel Table.

4 Results of Quantitative Assessment of AI Credit Scoring Transparency

The first step involves calculating an Aggregated Quantified Assessment (AQA) of the transparency level of an AI Credit Scoring tool. Participants propose numerical criteria individually for this assessment. In the proof-of-concept workshop, experts opted not to provide AQA or quantitative assessments of the current state of specific AI Credit Scoring (AI CS) due to concerns about the accuracy of such assessments. Although it was their first attempt to evaluate transparency, the experts expressed confidence in their ability to assess AI CS competencies. Instead of quantifiable data, they offered their individual intuitive opinions and insights. In the second step, the Quantitative Assessment of the Levels of Qualitative States (QALQS) is calculated based on Eqs. (3), (4), (5), (6), and (7). The results show a critically low state of 0.26, a low state of 0.36, a satisfactory state of 0.48, a good state of 0.61, and an excellent state of 0.75 (refer to Appendix D for a detailed calculation). These results form a generalized scale of states, as illustrated in Fig. 4.

Step three is a Qualitative Expert Judgement (QEJ) or Intuitive Survey Results (%) collected from 11 participants, resulting in a chart in Fig. 4. The survey question was “What do you think is the current state of AI Credit Scoring adherence to the ethical principle of transparency”? The majority of experts evaluated the state of AI Credit Scoring transparency as low (six votes). In contrast, two experts evaluated the state as satisfactory, two as critically low, and one participant evaluated it as a good state (Fig. 5).

Quantitative Assessment of the Average Collective Judgement of Experts (QAACJE), signifying the level of AI Credit Scoring transparency, is based on the results of expert judgments obtained from an intuitive survey and a matrix table. The process results in forming a new scale (E*S), which combines both qualitative and quantitative judgments of experts (see Fig. 4). For the tested group of experts, the average collective judgment resulted in a score of 0.38 (Fig. 6).

5 Findings, feasibility of the method, limitations, and outlook

The Expert Workshop (EW) serves as a valuable tool for conducting an in-depth analysis of the transparency level of AI-based credit scoring systems. Before the study, a clear definition and a proposed assumption of AI Credit Scoring’s transparency concept were established for the expert group. The expert group then had to confirm their understanding of AI ethics concepts and transparency. The following key steps were taken during the study:

1. Experts assessed the problem emotionally using the proposed scale from excellent to critically low. This was expressed quantitatively due to the methodology.
2. The experts identified and selected the five most informative characteristics, which served as a basis for establishing criteria to assess the transparency of AI-based credit scoring systems.
3. Characteristics enabled experts to designate the appropriate criteria levels for qualitative assessments of critically low, low, satisfactory, and excellent for each of the selected characteristics.
4. By applying the criteria selected by experts to indicate states of transparency, a generalized scale was derived considering all five indicators and their respective importance ratios.
5. The generalized scale enabled a quantitative assessment of a generalized qualitative judgment from a specific group of experts.

The collaborative efforts from the selected group of experts yielded results that expressed the opinions and judgments of this group. As per the QALQS developed by the participating group of experts, the consensus among experts regarding the adherence of AI Credit Scoring (AI CS) to the ethical principle of transparency falls within the range of “low” (0.36) to “satisfactory” (0.48); leaning more towards the “low” end of the spectrum.

The quantified assessment allows for easy comparison between different subjects or situations, which is particularly important when evaluating AI systems; especially in assessing ethical qualities. Moreover, if an Expert Workshop is conducted periodically, QAACJE allows for monitoring of how the transparency changes of AI CS over time. This gives evaluators and stakeholders full disclosure during the subject evaluation process, as they can see the numeric values and understand how the assessment was reached. The quantitative aspect of the workshop methodology serves as a competitive advantage, enabling constructive dialogue among experts by translating theoretical considerations into practical activities.

Additionally, the five unique characteristics formulated by the expert group, along with the initial fifteen characteristics proposed by three different groups during the second phase of the Expert Workshop (EW), hold significant importance. These characteristics shed light on perspectives that sometimes clash due to variations in the background knowledge of AI system implementations. Differences in the granularity of the initially formulated characteristics became apparent. Achieving consensus on characteristic definitions proved challenging due to disagreements among participants regarding the actors responsible for AI credit scoring transparency. Concerns were raised about the transferability of these characteristics to different jurisdictional layouts, which is a crucial aspect when regulating AI within responsible business contexts. Experts also noted the absence of a real-life AI Credit Scoring example for evaluation and identified a need for contextual settings in a use case description.

5.1 Feasibility of the method

The prior research on the landscape of AI ethical assessment frameworks identified their abundance[87] and, at the same time, the need for practical, quantifiable methods to evaluate AI adherence to ethical principles. The complexities associated with assessing AI ethicality were highlighted, especially considering the intricate nature of ethical concepts and the often opaque quality of existing assessments due to corporate confidentiality and non-disclosure agreements. In these terms, integrating various stakeholders' perspectives on ethical assessments were identified as critical aspects in addressing the practicability of ethical frameworks. Within this context, the feasibility of the Expert Workshop (EW) as a methodology to address these challenges was explored.

The study validated the feasibility of the EW methodology for assessing AI ethicality and shed light on the complexities and challenges involved in evaluating AI systems' ethicality. Moreover, it became evident that the specific conditions characterizing a concrete AI Credit Scoring tool depend on various factors, including the host company's corporate and business goals, industry conditions, market dynamics, and prevailing rules and regulations at a given point in time. This underscores the need for tailoring an Expert Workshop (EW) to the specific requirements of particular AI tools. Finally, the participation of experts who are deeply involved in developing specific AI systems and demonstrate strong motivation to ensure its competitive compliance with ethical AI standards is essential for conducting a high-quality assessment of such phenomena.

The Expert Workshop (EW) has demonstrated that its structured methodology provides valuable insights into the challenges and considerations associated with implementing responsible practices in business models utilizing AI. This conclusion is supported by the coherence observed across different results, emphasizing the methodology's potential effectiveness in evaluating the ethicality of AI systems. It provides a systematic way of developing quantifiable attributes to evaluate ethical compliance with the trustworthiness principles of AI systems.

5.2 Limitations

The limitations of the Expert Workshop method stem from the lack of comparable methods for research in expert elicitation methods. The metric system of weighted sums allows for statistical representation, however it still has the potential to produce biased results. Moreover, it is crucial to consider that psychological phenomena like group thinking might cause inaccurate or biased responses, and consequently yield biased quantified results. The requirement that characteristics need to be expressed as shares limits the evaluation to quantifiable aspects, which may overlook non-quantifiable ethical considerations. The technical and organizational challenges led to the fact that 10 participants out of thirteen could complete the matrix, and this aspect negatively affected the accuracy of the group result. Also, some limitations happened due to the discrepancy between the fast development of AI technologies and the lack of widespread certainty among experts in the field. This discrepancy was made apparent by experts who refused to provide judgment due to hesitation from a lack of concrete knowledge. As AI becomes increasingly implemented, there is a potential for this discrepancy to be bridged through expanded research.

5.3 Outlook

This paper has demonstrated that the assessment of AI Credit Scoring via the Expert Workshop (EW) can be achieved by obtaining quantified general estimates of the problem based on the expert opinion of a group. A set of measures is advisable to conduct a quality expert assessment, namely:

1. Improved data collection on characteristics could be achieved by expanding the pool of experts in AFI and soliciting expert opinions from a broader range of individuals. For example, further distribution of questionnaires and formed scales could help to collect more data from other experts who could not attend the current EW.[88]
2. Involving more types of stakeholders in the AI industry, such as policymakers, academics, start-up representatives, and public sector members, would be beneficial. Additionally, considering experts from other communities, cities, and countries would provide more diverse perspectives and enhance objectivity in problem-solving.
3. A given use case's "image" quality could be accomplished by increasing the number of characteristics considered, leading to a more comprehensive understanding of specific issues and potential solutions.
4. Testing other important AI ethical principles outlined by AI HLEG in the context of Expert Workshops (EW) could enable comparisons between ethical qualities to deepen understanding of the relationship between the ethical principles.

[87] Attard-Frost et al. [6].

[88] Pokholkov et al. [49].

5. Analyzing and comparing the results of multiple EWs could reveal similarities or differences in perceptions. Patterns that evolve from these comparisons could lead to the formation of a map that explores the perception of given use cases depending on a set of factors, such as stakeholder characteristics.

This approach could be instrumental in addressing the organizational challenges associated with implementing AI ethics. This is particularly relevant for rapidly evolving industries; where it can be challenging for self-identified experts to reach a consensus.

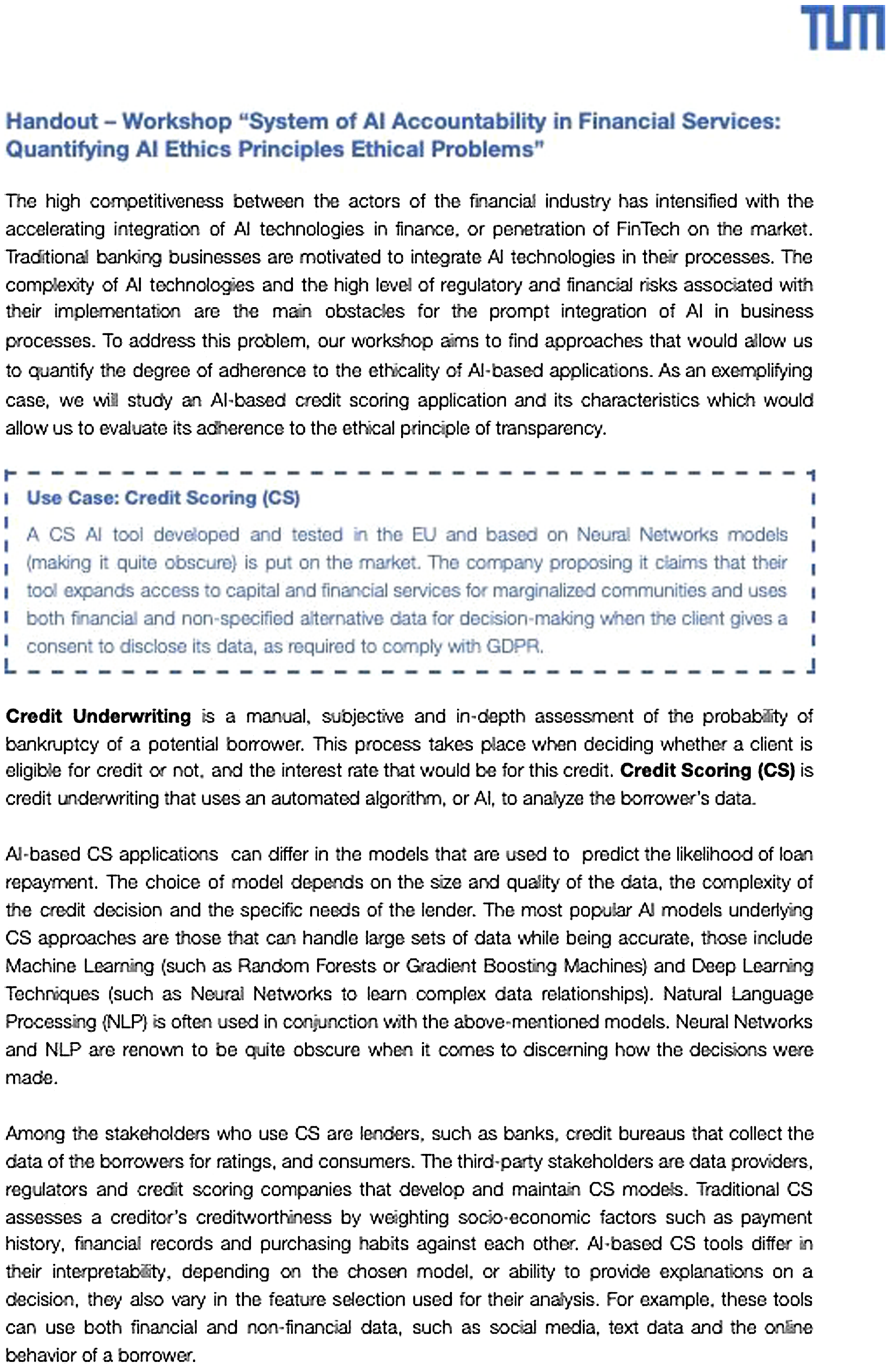
6 Conclusion

This study utilized the Expert Workshop (EW) methodology to define quantifiable adherence characteristics to ethical principles, focusing on transparency in AI-based credit scoring systems. Through a proof of concept EW, the study aimed to evaluate the effectiveness of the EW method in assessing the ethics of AI systems, particularly in the financial sector. Due to the type of expert elicitation method used, which provided relative estimates, numeric results were obtained through mathematical models. These results support the hypothesis that the Expert Workshop (EW) methodology is a viable approach for assessing the ethicality of AI systems.

Regarding the proof-of-concept results, experts' subjective opinions indicate low transparency achievement in AI CS technology. Experts provided a tentative scale for quantifying the adherence of such tools to the transparency principle and revealed the concerns about transparency in AI CS technology. Despite the workshop design nuances and the inherent subjectivity of the weighted sum metric, the study exemplifies the effectiveness of the methodology in this domain. All in all, the results of the initial stage demonstrate the exemplary study of EW methodology usage for assessing AI ethics components. In the meantime, there is an identified potential for evaluating a variety of ethical principles through the methodology of EW and assessing the comparative importance of principles in the context of AI. The proposed methodology can serve as a foundational framework for developing an ethical principles map, offering innovative insights into the ethical landscape of AI systems.

Appendix

Appendix A: Handout for Participants



TUM

Handout – Workshop "System of AI Accountability in Financial Services: Quantifying AI Ethics Principles Ethical Problems"

The high competitiveness between the actors of the financial industry has intensified with the accelerating integration of AI technologies in finance, or penetration of FinTech on the market. Traditional banking businesses are motivated to integrate AI technologies in their processes. The complexity of AI technologies and the high level of regulatory and financial risks associated with their implementation are the main obstacles for the prompt integration of AI in business processes. To address this problem, our workshop aims to find approaches that would allow us to quantify the degree of adherence to the ethicality of AI-based applications. As an exemplifying case, we will study an AI-based credit scoring application and its characteristics which would allow us to evaluate its adherence to the ethical principle of transparency.

Use Case: Credit Scoring (CS)

A CS AI tool developed and tested in the EU and based on Neural Networks models (making it quite obscure) is put on the market. The company proposing it claims that their tool expands access to capital and financial services for marginalized communities and uses both financial and non-specified alternative data for decision-making when the client gives a consent to disclose its data, as required to comply with GDPR.

Credit Underwriting is a manual, subjective and in-depth assessment of the probability of bankruptcy of a potential borrower. This process takes place when deciding whether a client is eligible for credit or not, and the interest rate that would be for this credit. **Credit Scoring (CS)** is credit underwriting that uses an automated algorithm, or AI, to analyze the borrower's data.

AI-based CS applications can differ in the models that are used to predict the likelihood of loan repayment. The choice of model depends on the size and quality of the data, the complexity of the credit decision and the specific needs of the lender. The most popular AI models underlying CS approaches are those that can handle large sets of data while being accurate, those include Machine Learning (such as Random Forests or Gradient Boosting Machines) and Deep Learning Techniques (such as Neural Networks to learn complex data relationships). Natural Language Processing (NLP) is often used in conjunction with the above-mentioned models. Neural Networks and NLP are renown to be quite obscure when it comes to discerning how the decisions were made.

Among the stakeholders who use CS are lenders, such as banks, credit bureaus that collect the data of the borrowers for ratings, and consumers. The third-party stakeholders are data providers, regulators and credit scoring companies that develop and maintain CS models. Traditional CS assesses a creditor's creditworthiness by weighting socio-economic factors such as payment history, financial records and purchasing habits against each other. AI-based CS tools differ in their interpretability, depending on the chosen model, or ability to provide explanations on a decision, they also vary in the feature selection used for their analysis. For example, these tools can use both financial and non-financial data, such as social media, text data and the online behavior of a borrower.

Appendix B: 15 Characteristics

Group 1

1. Share of relevant data points that were used in decision-making of AI CS that was
2. disclosed and explained to the customer.
3. Share of AI CS decisions that a credit analysis domain expert reviewed
4. Share of reviewed decisions by an AI CS, explanations on which were found satisfactory by a domain expert
5. Share of predictions correctly explained by a local interpretation method
6. Share of complaints/incidents asked on an AI CS decision after a customer asked for clarification on his/her decision

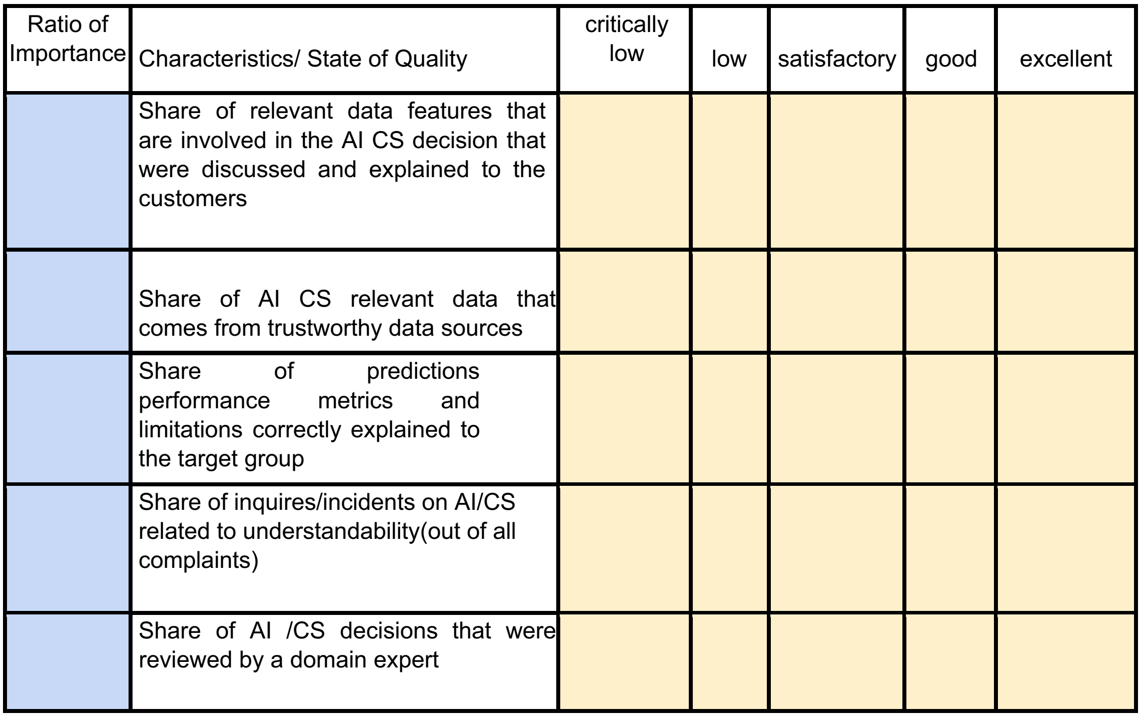
Group 2

1. weight of data source and type
2. share of cases where human intervention was needed
3. share of (sensitive) features used
4. model metrics (accuracy, confidence level, fairness metrics)
5. number of different data sources/share of trustworthy data sources

Group 3

1. Share documentation of relevant steps in the AI tool lifecycle (defined by standards and including post-hoc adjustments)
2. Share of cases for which output is reproducible within acceptable standards (defined by standards)
3. Share of group of users (reporting) understanding of the tool (UX research)
4. Share of known potential limitations presented to the public
5. Share of information about the system that is publically available (based on internal documentation)

Appendix C: Microsoft excel table



Ratio of Importance	Characteristics/ State of Quality	critically low	low	satisfactory	good	excellent
	Share of relevant data features that are involved in the AI CS decision that were discussed and explained to the customers					
	Share of AI CS relevant data that comes from trustworthy data sources					
	Share of predictions performance metrics and limitations correctly explained to the target group					
	Share of inquires/incidents on AI/CS related to understandability(out of all complaints)					
	Share of AI /CS decisions that were reviewed by a domain expert					

Appendix D: detailed calculations

The calculation of the Quantitative Assessment of the Levels of Qualitative States (QALQS) based on Equations (3), (4), (5), (6), and (7) is presented below.

1. Calculation of Step 2 (QALQS) with Eqs. (3), (4), (5), (6), (7).

critically low:

$(0.2 * 0.27)+(0.3 * 0.25) + (0.34 * 0.18) + (0.4 * 0.13) +(0.0 * 0.17)) = 0.2554–0.26;$

low:

$(0.4 * 0.27)+(0.5 * 0.25) + (0.43 * 0.18) + (0.5 * 0.13) +(0.1 * 0.17) = 0.3581–0.36;$

satisfactory:

$(0.5 * 0.27)+(0.6 * 0.25) + (0.52 * 0.18) + (0.7 * 0.13) +(0.1 * 0.17) = 0.4788–0.48;$

good:

$(0.6 * 0.27)+(0.7 * 0.25) + (0.62 * 0.18) + (0.8 * 0.13) +(0.2 * 0.17)=0.6136–0.61;$

excellent:

$(0.8 * 0.27)+(0.9 * 0.25) + (0.77 * 0.18) + (0.9 * 0.13) +(0.3 * 0.17) = 0.7528–0.75.$

Author contributions CL, EH, AB and MP contributed to the conception and design of the overall research project and the conception and planning of the conducted workshop. MP, EH, CL and AB contributed to the preparation, realization, post-processing and analysis of the workshop. Further, MP, AB, EH and CL contributed to writing, revising and approving the manuscript. All authors contributed to the article and approved the submitted version.

Funding Open Access funding enabled and organized by Projekt DEAL. This work was supported by Fujitsu Limited and the Technical University of Munich's Institute for Ethics in Artificial Intelligence (IEAI). All authors declare no other competing interests.

Data availability The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

Declarations

Conflict of interest The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Ethics statement Ethical review and approval were not required for the study on human participants in accordance with the local legislation and institutional requirements. Written informed consent for participation was not required for this study in accordance with the national legislation and the institutional requirements. Participants were made aware before participating in the workshop of the data collection and use for research purposes, as well as research questions and interests. All data have been anonymised in accord with the respect of participants' privacy.

References

1. Adams, S.J.: Projecting the next decade in safety management: a Delphi technique study. Prof. Saf.Saf. **46**(10), 26–29 (2001)
2. Ahmed, F.: Ethical aspects of artificial intelligence in banking. J. Res. Econ. Fin. Manage. **1**(2), 55–63 (2022). https://doi.org/10.56596/jrefm.v1i2.7
3. Amugongo, L.M., Bidwell, N.J., Corrigan, C.C.: Invigorating ubuntu ethics in AI for healthcare: enabling equitable care. In: Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency, Chicago, IL, USA, 12–15 June 2023, FAccT '23, pp. 583–592. Association for Computing Machinery, New York, NY, USA (2023)
4. Amugongo, L.M., Kriebitz, A., Boch, A., Lütge, C.: Operationalising AI ethics through the agile software development lifecycle: a case study of AI-enabled mobile health applications. AI Ethics (2023). https://doi.org/10.1007/s43681-023-00331-3
5. Andreotta, A.J., Kirkham, N., Rizzi, M.: AI, big data, and the future of consent. AI Soc. **37**(4), 1715–1728 (2022). https://doi.org/10.1007/s00146-021-01262-5
6. Attard-Frost, B., De los Ríos, A., Walters, D.R.: The ethics of AI business practices: a review of 47 AI ethics guidelines. AI Ethics **3**(2), 389–406 (2023). https://doi.org/10.1007/s43681-022-00156-6
7. Ayling, J., Chapman, A.: Putting AI ethics to work: are the tools fit for purpose? AI Ethics (2022). https://doi.org/10.1007/s43681-021-00084-x
8. Ben-David, A., Frank, E.: Accuracy of machine learning models versus "hand crafted" expert systems—a credit scoring case study. Expert Syst. Appl. **36**(3, Part 1), 5264–5271 (2009). https://doi.org/10.1016/j.eswa.2008.06.07
9. Butler, A.J., Thomas, M.K., Pintar, K.D.M.: Systematic review of expert elicitation methods as a tool for source attribution of enteric illness. Foodborne Pathog. Dis.Pathog. Dis. **12**(5), 367–382 (2015). https://doi.org/10.1089/fpd.2014.1844
10. Corrigan, C.C.: Lessons learned from co-governance approaches—developing effective AI policy in Europe. In: The 21 Yearbook of the Digital Ethics Lab, p. 2546. Springer International Publishing, Cham (2022)
11. Curto, G., Jojoa Acosta, M.F., Comim, F., Garcia-Zapirain, B.: Are AI systems biased against the poor? A machine learning analysis using Word2Vec and GloVe embeddings. AI Soc. (2022). https://doi.org/10.1007/s00146-022-01494-z
12. Dolganova, O.: Improving customer experience with artificial intelligence by adhering to ethical principles. Bus. Inform. **15**(2), 34–46 (2021). https://doi.org/10.17323/2587-814X.2021.2.34.46

13. Delbecq, A.L., Van de Ven, A.H., Gustafson, D.H.: Group Techniques for Program Planning. Scott, Foresman, and Co., Glenview, IL (1975)
14. Dwivedi, Y.K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P.V., Janssen, M., Jones, P., Kar, A.K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., et al.: Artificial intelligence (AI): multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. Int. J. Inf. Manage. **57**, 101994 (2021). https://doi.org/10.1016/j.ijinfomgt.2019.08.002
15. Eddy, Y.L., Bakar, E.M.N.E.A.: Credit scoring models: techniques and issues. J. Adv. Res. Bus. Manage. Stud. **7**(2), 2 (2017)
16. Executive Office of the President National Science and Technology Council: Preparing for the future of Artificial Intelligence [Online] (2016). https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf. Accessed 2 Mar 2024
17. Floridi, L.: Establishing the rules for building trustworthy AI. Nat. Mach. Intell. **1**(6), 261–262 (2019). https://doi.org/10.1038/s42256-019-0055-y
18. Fontes, C., Corrigan, C., Lütge, C.: Governing AI during a pandemic crisis: initiatives at the EU level. Technol. Soc. **72**, 102204 (2023)
19. Ghodselahi, A., Amirmadhi, A.: Application of artificial intelligence techniques for credit risk evaluation. Int. J. Model. Optim. (2011). https://doi.org/10.7763/IJMO.2011.V1.43
20. Garthwaite, P.H., Kadane, J.B., O'Hagan, A.: Statistical methods for eliciting probability distributions. J. Am. Stat. Assoc. **100**(470), 680–701 (2005). https://doi.org/10.1198/016214505000000105
21. Hallensleben, S., Fetic, L., Fleischer, T., Grünke, Hagendorff, T., Hauer, M., Hauschke, A., Heesen, J., Herrmann, M., Hillerbrand, R., Hubig, C., Kaminski, A., Krafft, T., Loh, W., Otto, P., Puntschuh, M., Hustedt, C. (2020). From Principles to Practice: An Interdisciplinary Framework to Operationalize AI Ethics. https://publikationen.bibliothek.kit.edu/1000121427
22. Hagendorff, T.: AI virtues—the missing link in putting AI ethics into practice. Philos. Technol. **35**(3), 55 (2022). https://doi.org/10.1007/s13347-022-00553-z
23. Hagendorff, T.: The ethics of AI ethics: an evaluation of guidelines. Mind. Mach. **30**(1), 99–120 (2020). https://doi.org/10.1007/s11023-020-09517-8
24. HLEG: Ethics guidelines for Trustworthy AI Shaping Europe's digital future [Online] (2019). https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai. Accessed 2 Mar 2024
25. Hohma, E., Boch, A., Trauth, R., Lütge, C.: Investigating accountability for Artificial Intelligence through risk governance: a workshop-based exploratory study. Front. Psychol. **14**, 1073686 (2023)
26. Hooks, D., Davis, Z., Agrawal, V., Li, Z.: Exploring factors influencing technology adoption rate at the macro level: a predictive model. Technol. Soc. **68**, 101826 (2022). https://doi.org/10.1016/j.techsoc.2021.101826
27. Hsu, C.C.: The Delphi technique: making sense of consensus. Pract. Assessment Res. Eval. **12**(1), 1–8 (2007)
28. Hsu, C.-C., Sandford, B.A.: The Delphi technique: making sense of consensus. Pract. Assess. Res. Eval. **12**, 10 (2019). https://doi.org/10.7275/pdz9-th90
29. Huang, Z., Chen, H., Hsu, C.-J., Chen, W.-H., Wu, S.: Credit rating analysis with support vector machines and neural networks: a market comparative study. Decis. Support. Syst.. Support. Syst. **37**(4), 543–558 (2004). https://doi.org/10.1016/S0167-9236(03)00086-1
30. IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems: Ethically Aligned Design [Online] (2019). https://standards.ieee.org/industry-connections/ec/autonomous-systems.html. Accessed 2 Mar 2024
31. Jammalamadaka, K.R., Itapu, S.: Responsible AI in automated credit scoring systems. AI Ethics **3**(2), 485–495 (2023). https://doi.org/10.1007/s43681-022-00175-3
32. Koefer, F., Lemken, I., & Pauls, J. (2023). Realizing fair outcomes from algorithm- enabled decision systems: an exploratory case study. In: Lecture Notes in Business Information Processing, vol. 467 LNBIP, pp. 52–67. Scopus. https://doi.org/10.1007/978-3-031-31671-5_4
33. Van Lenthe, J.: ELI: an interactive elicitation technique for subjective probability distributions. Organ. Behav. Hum. Decis. Process. Behav. Hum. Decis. Process. **55**(3), 379–413 (1993). https://doi.org/10.1006/obhd.1993.1037
34. Jobin, A., Ienca, M., Vayena, E.: The global landscape of AI ethics guidelines. Nat. Mach. Intell. **1**(9), 389–399 (2019). https://doi.org/10.1038/s42256-019-0088-2
35. Koh, H.-K., Burnasheva, R., Suh, Y.G.: Perceived ESG (environmental, social, governance) and consumers' responses: the mediating role of brand credibility, brand image and perceived quality. Sustainability (2022). https://doi.org/10.3390/su14084515
36. Kozodoi, N., Jacob, J., Lessmann, S.: Fairness in credit scoring: assessment, implementation and profit implications. Eur. J. Oper. Res.Oper. Res. **297**(3), 1083–1094 (2022). https://doi.org/10.1016/j.ejor.2021.06.023
37. Kumar, A., Sharma, S., Mahdavi, M.: Machine learning (ML) technologies for digital credit scoring in rural finance: a literature review. Risks **9**(11), 192 (2021). https://doi.org/10.3390/risks9110192
38. Maree, C., Modal, J. E., & Omlin, C. W. (2020). Towards responsible AI for financial transactions. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp 16–21. https://doi.org/10.1109/SSCI47803.2020.9308456
39. Max, R., Kriebitz, A., VonWebsky, C.: Ethical considerations about the implications of artificial intelligence in finance. In: San-Jose, L., Retolaza, J.L., van Liedekerke, L. (eds.) Handbook on Ethics in Finance, pp. 577–592. Springer International Publishing, Cham (2021)
40. Mills, S.: The misuse of algorithms in society (SSRN Scholarly Paper 4400026). SSRN J. (2023). https://doi.org/10.2139/ssrn.4400026
41. Morandín-Ahuerma, F. (2023). Montreal Declaration for Responsible AI: 10 Principles and 59 Recommendations. OSF Preprints. https://doi.org/10.31219/osf.io/sj2z5
42. Morgan, M.G.: Use (and abuse) of expert elicitation in support of decision-making for public policy. Proc. Natl. Acad. Sci. **111**(20), 7176–7184 (2014). https://doi.org/10.1073/pnas.1319946111
43. Morley, J., Kinsey, L., Elhalal, A., Garcia, F., Ziosi, M., Floridi, L.: Operationalising AI ethics: barriers, enablers and next steps. AI Soc. **38**(1), 411–423 (2023). https://doi.org/10.1007/s00146-021-01308-8
44. Morrison-Saunders, A., Retief, F.: Walking the sustainability assessment talk—progressing the practice of environmental impact assessment (EIA). Environ. Impact Assess. Rev. **36**, 34–41 (2012). https://doi.org/10.1016/j.eiar.2012.04.001
45. Nowakowski, M., Waliszewski, K.: Ethics of artificial intelligence in the financial sector. Przegląd Ustawodawstwa Gospodarczego **2022**, 2–9 (2022). https://doi.org/10.33226/0137-5490.2022.1.1
46. Omrani, N., Rivieccio, G., Fiore, U., Schiavone, F., Agreda, S.G.: To trust or not to trust? An assessment of trust in AI-based systems: concerns, ethics and contexts. Technol. Forecast. Soc. Chang. **181**, 121763 (2022). https://doi.org/10.1016/j.techfore.2022.121763
47. Ozili, P.K.: Big data and artificial intelligence for financial inclusion: benefits and issues (SSRN Scholarly Paper 3766097). SSRN J. (2021). https://doi.org/10.2139/ssrn.3766097

48. Pasman, H.J., Rogers, W.J.: How to treat expert judgment? With certainty it contains uncertainty! J. Loss Prev. Process Ind. **66**, 104200 (2020). https://doi.org/10.1016/j.jlp.2020.104200
49. Pokholkov, Y., Horvat, M., Quadrado, J.C., Chervach, M., Zaitseva, K. (2020). Approaches to assessing the level of engineering students' sustainable development mindset. In: 2020 IEEE Global Engineering Education Conference (EDUCON), pp. 1102–1109. https://doi.org/10.1109/EDUCON45650.2020.9125292
50. Podinovski, V., Potapov, M.: Weighted sum method in the analysis of multicriterial decisions: pro et contra. Bus. Inf. **3**(25), 41–48 (2013)
51. Radclyffe, C., Ribeiro, M., Wortham, R.H.: The assessment list for trustworthy artificial intelligence: A review and recommendations. Front. Artif. Intell. (2023). https://doi.org/10.3389/frai.2023.1020592
52. Satybaldiyeva, E., et al.: Applying the export method to determine a company. Transp. Probl. **18**(2), 123–132 (2023). https://doi.org/10.20858/tp.2023.18.2.11
53. Savinova, O.V.: Probation of an expert seminar on "students' involvement in research work during studying. Inzhener Obrazov **29**, 34–44 (2021). https://doi.org/10.4835/18102883_2021_29_3
54. Saaty, R.W.: The analytic hierarchy process—what it is and how it is used. Math. Modell. **9**(3), 161–176 (1987). https://doi.org/10.1016/0270-0255(87)90473-8
55. Solanki, R.: Fintech: a disruptive innovation of the 21st century, or is it? Glob. Bus. Manage. Res. **14**(2), 76–87 (2022)
56. Solove, D.J.: A taxonomy of privacy. Univ. Pa. Law Rev. **154**(3), 477 (2006). https://doi.org/10.2307/40041279
57. Spiekermann, S., Winkler, T.: Value-based engineering for ethics by design. SSRN Electron. J. (2020). https://doi.org/10.2139/ssrn.3598911
58. Stix, C.: Actionable principles for artificial intelligence policy: three pathways. Sci. Eng. Ethics **27**(1), 15 (2021). https://doi.org/10.1007/s11948-020-00277-3
59. Szwed, P.: Working Paper: Establishing a Theoretically Sound Baseline for Expert Judgment in Project Management – Part I. [Online] (2014). https://www.researchgate.net/publication/259948022_Working_Paper_Establishing_a_Theoretically_Sound_Baseline_for_Expert_Judgment_in_Project_Management_-_Part_I. Accessed 2 Mar 2024
60. Tolkacheva, K. (2015). Expert seminar as a form of realizing the goals of problem-oriented training of specialists in engineering and technology. National Research Tomsk Polytechnic University (TPU). Retrieved from http://catalog.lib.tpu.ru/catalogue/simple/document/RU/TPU/book/336904
61. Tripopsakul, S., Puriwat, W.: Understanding the impact of ESG on brand trust and customer engagement. J. Hum. Earth Future **3**(4), 430–440 (2022). https://doi.org/10.8991/HEF-2022-03-04-03
62. Truby, J., Brown, R., Dahdal, A.: Banking on AI: mandating a proactive approach to AI regulation in the financial sector. Law Fin. Markets Rev. **14**(2), 110–120 (2020). https://doi.org/10.1080/17521440.2020.1760454
63. TUV SÜD: Artificial Intelligence. [Online] (2023). https://www.tuvsud.com/en/themes/artificial-intelligence. Accessed 2 Mar 2024
64. UK Parliament Committee: Written Evidence Submitted by Committee on Standards in Public Life (GAI0110). [Online] (2022). https://committees.parliament.uk/writtenevidence/114057/html/. Accessed 2 Mar 2024
65. Usher, W., Strachan, N.: An expert elicitation of climate, energy, and economic uncertainties. Energy Policy **61**, 811–821 (2013). https://doi.org/10.1016/j.enpol.2013.06.110
66. Vakkuri, V., Kemell, K.-K., Jantunen, M., Halme, E., Abrahamsson, P.: ECCOLA—a method for implementing ethically aligned AI systems. J. Syst. Softw.Softw. **182**, 111067 (2021). https://doi.org/10.1016/j.jss.2021.111067
67. Wachter, S., Mittelstadt, B., Russell, C.: Why fairness cannot be automated: bridging the gap between EU non-discrimination law and AI. Comput. Law Secur. Rev.. Law Secur. Rev. **41**, 105567 (2021). https://doi.org/10.1016/j.clsr.2021.105567
68. Winfield, A.F.T., Jirotka, M.: Ethical governance is essential to building trust in robotics and artificial intelligence systems. Philos. Trans. R. Soc. A Math. Phys. Eng. Sci. **376**(2133), 20180085 (2018). https://doi.org/10.1098/rsta.2018.0085
69. Zhou, J., Chen, F., Berry, A., Reed, M., Zhang, S., Savage, S. (2020). A survey on ethical principles of AI and implementations. In: 2020 IEEE Symposium Series on Computational Intelligence (SSCI), pp. 3010–3017. https://doi.org/10.1109/SSCI47803.2020.9308437