Crediting Humans: A Systematic Assessment of Influencing Factors for Human-in-the-Loop Figurations in Consumer Credit Lending Decisions

Theresa Züger
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
zueger@hiig.de

Philipp Mahlow
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
philipp.mahlow@hiig.de

Daniel Pothmann
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
daniel.pothmann@hiig.de

Katharina Mosene
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
katharina.mosene@hiig.de

Fabian Burmeister
University of Hamburg
Hamburg, Germany
fabian.burmeister@uni-hamburg.de

Matthias Kettemann
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
matthias.kettemann@hiig.de

Wolfgang Schulz
Alexander von Humboldt Institute for Internet and Society
Berlin, Germany
wolfgang.schulz@hiig.de

Abstract

The concept of 'human-in-the-loop' (Hilo) has gained prominence as a regulatory mechanism for ensuring human control in automated systems, particularly in the context of automated decision-making (ADM) mechanisms or AI-supported systems. However, despite its increasing use in regulatory discourse, there is a lack of empirical understanding regarding the real-world conditions and influencing factors that affect decision-making processes in such hybrid systems. This paper aims to address this gap by focusing on the use of automation, and the inclusion of humans, in consumer-facing credit lending decisions, a key industry use case. By employing an interdisciplinary approach that combines legal perspectives, social sciences, and architectural modeling—a methodology rooted in computer science—this research offers a first systematic analysis of the factors that influence meaningful human control in

(semi-)automated decision processes. Specifically, it contributes to the field by broadening the understanding of the Hilo concept and providing empirical insights into the factors that influence human-machine interaction in co-decisionary architectures. Referencing existing literature that proposes to differentiate various roles for Hilos, the paper proposes two additional roles of Hilos to be considered, due to its empirical findings, namely a special case handling role and a broader understanding of the resilience role of Hilos.

FAccT '25, Athens, Greece

ACM ISBN 979-8-4007-1482-5/2025/06
https://doi.org/10.1145/3715275.3732086

CCS Concepts

• **Applied computing**; • **Law, social and behavioral sciences**; • **Law**;

Keywords

human-in-the-loop (Hilo), consumer credit lending, automated decision-making, meaningful human control, AI, HITL

ACM Reference Format:
Theresa Züger, Philipp Mahlow, Daniel Pothmann, Katharina Mosene, Fabian Burmeister, Matthias Kettemann, and Wolfgang Schulz. 2025. Crediting Humans: A Systematic Assessment of Influencing Factors for Human-in-the-Loop Figurations in Consumer Credit Lending Decisions. In *The 2025 ACM Conference on Fairness, Accountability, and Transparency (FAccT '25), June 23–26, 2025, Athens, Greece*. ACM, New York, NY, USA, 12 pages. https://doi.org/10.1145/3715275.3732086

1 INTRODUCTION

The human-in-the-loop (Hilo) concept is increasingly used as a regulatory tool to initiate and mandate meaningful human control in automated systems. Hilos are implemented in automated decision-making (ADM) processes to fulfill a number of roles, such as correcting a system recommendation (corrective role) or helping users interact with the system (interface role) [1]. Importantly, they are also employed to assign accountability and liability to [2, 3]. However, Hilos should not have to assume liability without being provided with adequate measures and conditions to meaningfully influence the decision outcome. It is therefore necessary to understand which influencing factors affect human-machine interactions in what way, so that meaningful human agency is enabled. While it has been widely discussed in the regulatory discourse on automation, especially since the increasing relevance of artificial intelligence (AI) systems, the concept is hardly ever based on a deeper empirical understanding of such conditions and influencing factors for actual humans in the loop in real-world settings. As a result, there is a lack of understanding of the influencing factors that humans in the loop of ADM processes might face. Our research aims to fill this knowledge gap and thereby contribute to the field of influencing factors in fairness, accountability and transparency, while focusing on an industrial use case, namely the automation of credit lending. One the one hand, the influencing factors can be used to evaluate existing or planned Hilo implementations for their potential to have a meaningful impact on the decision-making process. Conversely, they can be used in the regulatory discourse, which this paper only makes reference to, to determine the quality of requirements for human intervention in automated processes. We followed an interdisciplinary approach, combining legal perspectives with social sciences, as well as a methodological approach called architectural modelling, which is rooted in computer science. Only this multifaceted approach allowed us to systematically analyze the influencing factors for meaningful human control in automation. This research is part of an ongoing research effort with the overall goal of developing a guiding taxonomy of such factors. However, this particular paper will zoom in on the first iteration and focus on insights into relevant influence factors for the case of credit lending.

The main contributions of this paper are:

1. Reflecting on existing conceptualizations of the Hilo and proposing a broader definition.
2. Enriching the concept of the Hilo with empirical evidence on relevant influencing factors for Hilos in the case of credit lending.
3. Introducing additional roles to the Hilo discourse, namely a special case handling role, and arguing for a broader understanding of the resilience role of Hilos.

2 BACKGROUND

This section begins with an overview and reflection of literature on the concept of the 'human-in-the-loop' in decision-making processes. Following this reflection, it introduces our definition and explains the lending process together with its legal background.

'Human-in-the-loop' is a technical term that originated in the field of computer science and describes controls in the machine learning cycle to improve the accuracy of models [4]. This idea, together with its label, has by now entered several non-technical disciplines, notably law [5–7], in the context of the regulation of (semi-)automated decision-making processes. The term often provokes ambiguous interpretations as to which human and which loop specifically is targeted. For example, Article 22 of the EU General Data Protection Regulation (GDPR) prohibits (under certain circumstances) subjecting individuals to purely automated decisions, which has led to the pro forma inclusion of humans in increasingly automated decision-making processes. According to current conceptualizations in the literature—especially in the field of law—Hilos are supposed to ensure accountability, preserve human agency and fulfill constitutional principles (including dignity), among other values and factors associated with the inclusion of humans [3, 6, 8–10]. Looking at some of the practical contexts in which the Hilo concept is being discussed, the challenge of maintaining ethical decision making and individual justice versus operability in the face of sheer volume becomes apparent [8]. This problem has been described as a "fairness-utility trade-off" [11]. In addition to ensuring 'due process', public trust in a decision outcome is significantly influenced by the actor making the decision, with higher initial trust achieved when humans are 'in the loop' than when decisions are made by fully automated systems [12, 13]. This effect justifies a dual focus on both the accuracy of systems and public acceptance of procedures and outcomes [14].

Considering existing and planned legal documents containing Hilo requirements, it is tempting to adopt the regulatory assumption that the inclusion of 'a human' in 'the loop' ensures (more) correct and (more) ethical decision-making. However, critical voices in the literature have raised concerns over the lack of adequate consideration of the individual's capacity to meaningfully contribute to or intervene in the decision-making process [1, 3, 10]. Crucially, human capabilities and individual factors (including risk aversion, internal goal conflicts, framing effects, situational effects, "workload, situational awareness, stress and fatigue" [15, 17, 18]) rarely receive adequate attention. Not only do these factors affect human judgement and decision-making in general, but their specific combination leads to different behavioral manifestations depending on the individual [16]. For example, stress is perceived and responded to differently by different people. If a person's involvement is to be meaningful in terms of fulfilling the above expectations, psychological variables should be considered in any process and interface design.

Another criticism of the current understanding of the Hilo concept concerns the assumption of a binary liability construct [2]. This critique argues that, from a regulatory perspective, the inclusion of a Hilo exempts a company that uses an ADM system for most of the decision process from the restrictions imposed on fully automated decision making, such as Art. 22 GDPR, regardless of whether the Hilo has the ability to meaningfully participate in the decision or merely 'rubber stamps' the system's recommendations. Such a rubber stamping is especially problematic, where Hilos are included into an ADM process specifically so that liability can be attributed to them [2, 3]. It is therefore argued that further discussion is needed on what criteria ensure meaningful human intervention, so that the Hilo can actually review the results provided by the ADM systems. The European Court of Justice (ECJ) underlined this in its much-publicized so-called "Schufa decision" (Case C-634/21). The ECJ found that a decision-making process is "based solely on automated processing" within the meaning of Art. 22 (1) GDPR, even if a human actor is added to the decision-making loop, as long as the human decision "draws strongly" on the output provided by the ADM system (Recital 73). In other words: A decision process design in which a human actor merely accepts a recommendation from an ADM system without further regard is still considered an automated decision.

This shows that the Hilo concept is inextricably linked to the notion of meaningful human control (MHC), which includes the maintenance and reinforcement of "individual human control and moral responsibility within a complex (technical) system" [19]. The following relationship is proposed: any Hilo should be able to exercise MHC. However, it has been posited that not every instance of MHC needs to be executed by a human actor, as it has been suggested that human-like control could be considered sufficient even when executed by non-human entities [20, 21]. As noted by De Sio and van den Hoven, this is the case when a decision-making system is consistent with human moral reasoning.

To add to the complexity, many, if not most, real-world decisions in organizational contexts are unlikely to be made in isolation by a single actor. Accordingly, this increases the risk of misalignment of values between strategic decision makers (e.g., boards of directors) and practical decision makers (e.g., front-desk advisors) with unmet information needs due to context-specific, pragmatic challenges in implementing these values [10, 16]. As (semi-)automated systems are increasingly integrated into such processes, issues of non-transparency could be catalyzed, as certain steps are no longer instructed to be carried out by a human, but opaquely automated. This opacity reveals the intertwined fallacy of the common expectation to simply combine the strengths of humans (e.g., reasoning or context awareness) and machines (e.g., speed or consistency). This conclusion is short-sighted, as their interaction brings about its own strengths and weaknesses, including risks such as automation bias as well as the progressive fading of skills [1], which negatively affects the situational and long-term supervisory effectiveness of the individual. It is therefore essential to mediate these issues with longer-term observations and empirically tested measures such as careful, human-centered interface design with a focus on intuitive usability [22]. From a more strategic perspective, the structural incompatibility of machine and human decision making raises procedural questions about efficient integration [23], many of which remain to be answered. Some of these questions include how to establish complementary human-machine interaction and what effective 'explainability' of AI systems might look like [24, 25].

In her comprehensive work on the legal conceptualization of the human-in-the-loop, Crootof [1] describes nine non-exclusive roles humans can take in automated processes, which were labeled as follows: corrective, resilience, justificatory, dignitary, accountability, 'stand-in', friction, 'warm body', and interface. A full appraisal of the publication is highly recommended for a detailed description of every role. However, it shall be noted that due to imprecision in legal requirements as well as in occupational instructions, it is unlikely that said roles can be fulfilled in practice. Crootof explicitly criticizes how decisions are often viewed as isolated situations where humans are put in place to absorb liability and serve as a moral crumping zone, without sufficient training for and explanation of their intended function in the ADM-process. According to her, the key "question for successful automation is not who has control over what or how much—it is how do we get along together?" [1].

Given the practical complexity of human-machine 'collaborative' decision-making in organizational contexts, it is crucial to fully consider the relevant factors that influence the realization of a decision. This requires reflection on the entire human-led process—as part of a socio-technical system—leading up to the subsequent practical situation(s), as the latter seems to be a key concern of most Hilo regulations [1]. Only a careful, system-wide integration of these factors in the setting of normative frameworks can fulfill the promise of the intended fairness, accountability and transparency of governance measures.

2.1 Defining 'human-in-the-loop'

Different disciplines have made important contributions to the understanding of the human-in-the-loop concept, but there is no cross-disciplinary and generally accepted definition. On the contrary, it is striking that different disciplines have different understandings of the Hilo. In particular, the approaches differ with regard to the stage in the life cycle of an automated system at which the human actor considered to be a Hilo interacts with the system.

Publications from the computer science literature predominantly understand the Hilo as a human actor who interacts with AI systems in their development and training phase in order to combine human and machine 'intelligence' to increase the accuracy of the trained systems [4]. It can be said that a coherent understanding of the Hilo concept has emerged in computer science [26]. However, such a narrow understanding, which focuses only on the development process of systems, does not allow for a comprehensive assessment of the interaction between humans and machines as part of (semi-)automated processes within complex socio-technical systems [1]. Such a definition does not capture interactions in the context of operating a hybrid decision system. The different roles of human influence on machine systems cannot be adequately analyzed in this way.

Regulatory definitions, on the other hand, specifically take into account the human influence on a deployed decision-making system during its operation. The European Commission, for example, understands human-in-the-loop as "human intervention in every decision cycle of the system" [27]. The Commission then concludes that such an implementation of human involvement would be 'neither possible nor desirable' and that this concept is therefore not recommended for the implementation of human oversight. However, despite this criticism of the Hilo concept, relevant regulatory publications continue to refer to and reflect on this regulatory understanding [1]. Crootof et al. [1] acknowledge this working definition, but at the same time criticize it as narrow, since it only addresses the systems implementation phase. They therefore suggest that regulators should not overlook other significant human contributions to a decision-making process, for instance, in earlier phases of an AI-lifecycle.

The present paper further supports Crootof et al.'s criticism of the limited regulatory understanding of the Hilo concept. This narrow understanding ignores the fact that human influence on many stages and different loops of a decision system can significantly affect the quality of decisions. Therefore, the Hilo definition should include interactions during iterative revision cycles and not be limited to humans interacting with the system at each individual decision cycle. For this reason, we propose a mediating and broader definition of the Hilo concept that includes the production and improvement process of decision systems, as well as upstream and downstream processes and the actual deployment phase. In all of these phases, human influence can have a significant impact on the results or decisions of the overall system, depending on the degree of actual involvement. The definition includes (semi-)automated decision-making systems and not only those systems that are primarily discussed under the ambiguous term AI. Therefore, the term "automated decision-making systems", as used, for example, in Art. 22 GDPR is utilized instead.

We propose a broader definition of the human-in-the-loop concept as an umbrella term to describe the roles and interactions of human actors in automated decision-making systems throughout their lifecycle, which have a significant impact on the quality of decision-making.

This definition allows for a more holistic approach to identifying actors as Hilo who have a significant impact, rather than focusing solely on the development process or the deployment phase. The quality of a decision is a context-dependent measure that can be constructed from a variety of dimensions, including the perspective of the actors involved. In the case of credit institutions, a good credit decision is defined as one that is at least profitable, has low default rates, and is legally compliant. Beyond these indicators, further standards—which may be of a legal or ethical nature—may be relevant for assessing the quality of a decision, such as non-discrimination. Discrimination is defined here as any unjustified (direct or indirect) disadvantage in the sense of the principles of equality in Article 3 of the German Constitution and Section 3 of the German Equal Treatment Act (AGG). Important elements of decision quality from a philosophical perspective, proposed by Santoni de Sio and van den Hoven [21], are that decisions follow human and moral reasons as well as the legal framework. The absence of arbitrariness, the reliance on rational argumentation, and the grounding of the decision in the "well-founded general ideas of justice of the community" [28] are also seen as further indicators of the quality of a decision.

The limitation of the Hilo definition to human actors with significant influence serves to limit the scope of this definition. Consequently, only incidentally relevant contributions to the final decision should be excluded. In order to determine the significance of a contribution, it is necessary to consider it in the context of the specific individual process. Therefore, the definition of general criteria for its existence must be made with reservations. In any case, it is likely to be relevant to consider the proximity of a contribution to the decision in question and the agency of the human actor in question in terms of their ability to influence the decision.

We recognize, however, that the degree of significant influence is open to debate. This need for context and interpretation is productive in the sense that it requires reflection on the identification of the relevant Hilos in each specific use case.

2.2 Methods

The guiding research question of our work is: Which influencing factors have a significant impact on the human-in-the-loop and the quality of decision-making from an applied perspective? We approached this question in the following way.

Our study followed an interdisciplinary research approach involving scholars from law, social sciences, and computer science, which allowed us to examine the factors influencing Hilos from multiple disciplinary perspectives in an exploratory manner. Specifically, our research approach built on several existing methodologies and combined them for this purpose. At its core, we adapted Burmeister et al.'s [29] ARBITER method ("architecture-based modeling method for ecosystem-based regulation"), which supports the iterative and systematic exploration of complex socio-technical ecosystems through architectural modeling, aiming to close the discursive gap between information science and law. Starting from their abstract meta-model as a framework, our approach included structurally decomposing the decision-making process into its elemental 'blocks' and mapping them on horizontal (functional separation) and vertical (organizational separation) layers, comprising relevant human actors with explicit and implicit roles as well as technological systems with defined purposes. The inclusion of distinctive human actors at spatiotemporally separated stages of the process in particular allowed for a more holistic integration of co-dependent organizational components with variable functions, interests, and incentives, as well as their reciprocal relationships in a coherent, examinable visualization. By applying this method and adjusting it to our research question, we were able to understand the decision-making ecosystem for the field of credit lending in its full complexity, i.e., the socio-technical interplay of actors, systems, processes, and interfaces beyond mere data flows. In order to study the lending case addressed in this paper, we conducted three iterations of data collection (desk research, roundtable discussions, and expert interviews), architecture model creation and refinement according to the ARBITER method [29], and derivation of relevant influencing factors on a micro- and macro-level through a PESTEL analysis [30]. The PESTEL analysis is a structured, heuristic approach from economics to identify political, economic, social, environmental, and legal factors that potentially affect the analyzed issue to be decided [30]. In the first iteration, we conducted a superficial, exploratory research based on web searches and unstructured literature reviews to gain an initial overview of the credit lending ecosystem and define an adequate level of model abstraction. For our search, we focused on the keywords "human-in-the-loop" (multiple versions of the term), "automated decision making", and "automated lending", via the database "Primo ExLibris web library, University of Liverpool".

Highly relevant sources for our study were previous works by Crootof et al. [1] and Tschiatschek et al. [10]. We analyzed these sources with the aim of extracting frequently mentioned socio-technical elements and relationships that were considered relevant to credit lending. Based on these initial findings, we created a primitive architecture model that outlined the credit and lending ecosystem. For example, key elements included the most relevant actors, such as the consumer and the banker, and their relationships during the lending process. However, these initial findings were considered insufficient to derive fundamental drivers for Hilo and lacked empirical support.

Therefore, in a second iteration, we conducted a roundtable discussion with a group of 13 experts from different fields, some from banks, credit bureaus, researchers in the field, consumer advocates, as well as relevant NGOs from the financial and technology sector. The roundtable discussion was recorded via anonymized note-taking and subsequently coded in a qualitative content analysis [31]. To do this, we created an initial coding scheme to guide our analysis. The coding scheme included deductive codes from previous research, such as "actors", "process information", and the various categories of a PESTEL analysis, and was refined with emerging codes during the coding process.

Based on the coded data, we refined our architecture model to provide a more nuanced visualization of the lending ecosystem. For example, we added elements such as a more detailed understanding of the actors involved, including the front-desk advisor, the risk analyst, management, or the audit team, reflecting their functional relationships, dependencies, and responsibilities. We used both the architecture model and a PESTEL analysis to derive the first set of influencing factors for Hilos. Specifically, we used PESTEL analysis to focus and structure the influencing factors that we derived from the deeper understanding of the lending process from our architecture model. However, it became clear that more 'insider information' was needed to better understand the actual drivers, which are often implicit.

Therefore, in a third iteration, we conducted 13 semi-structured interviews with a total of 19 experts (in some interviews we talked to more than one expert from different departments of one institution). We selected the interviewees based on their extensive experience and expertise in credit lending and their diverse positions in different organizations. Ten of these experts were employees of lending institutions in various positions, ranging from management to consulting, auditing, or IT. The other nine experts hold positions in various institutions relevant to our context, including consumer credit counseling, debt counseling, financial research, law, representatives from Schufa (a German credit referencing agency), and representatives from BaFin (the German Federal Financial Supervisory Authority). All interviews were conducted following a responsive interview approach [32], and the interview guide was developed following Helfferich [33], consisting of narrative prompts for each section as well as optional follow-up questions. In detail, the interview questions were tailored to the specific expert or group of experts each time, but recurring key questions related to the interviewees' workflows, their understanding of the technological systems they utilize, and their perceptions of the role of automation in lending. All interviews were transcribed and coded using MAXQDA and subjected to further qualitative content analysis [31] with the specific aim of identifying new information for refining the architecture model and evidence of influencing factors. We reused our coding scheme from the second iteration and further refined it during the coding process. Our qualitative content analysis consisted of two cycles of coding. First, following the coding scheme, we coded all text passages throughout the transcripts, resulting in 694 codes. Second, using axial coding, we aggregated the codes into overarching categories. For example, major categories were "decision-making process", "influencing factors", and "meta-learning". The analyzed data served to identify additional socio-technical elements and relationships, and helped us fill in gaps, correct assumptions, and highlight neuralgic points in the architecture model to gain a more coherent understanding of the lending process and the various roles of Hilos within this complex ecosystem. Finally, we conducted another PESTEL analysis to identify other relevant influencing factors.

In a meta-analysis, following an extractive and structured approach [34, 35] we reflected on the structured results of the PESTEL analysis with reference to our process knowledge from the architecture model to identify which are the most relevant influencing factors for the Hilo as an abstract concept.

2.3 Case Study Description

We focused our research to provide insights on empirical factors influencing humans in the loop in the case of credit lending for consumer loans with a limit of about one million Euros, which marks a common risk-barrier in the industry. Evaluation above one million Euros requires even more involvement of humans, according to bank internal experts. The scope was limited to the German context.

The process of obtaining a loan is initiated by the applicant submitting an application to a designated credit institution, which can be done either online or in person at a physical branch. It is also possible to engage the services of an advisor who may act independently, on behalf of or in cooperation with a credit institution. When an application is submitted, the bank initiates an analysis based on both internal and external data. It is important to note that internally, a rating based on existing bank data (e.g., income and expenses, as well as the history of bank account transactions) is primarily used, while externally, a so-called 'credit score' is provided, which is made up of various key figures (e.g., customer payment defaults, number of credit cards, number of current loans). The credit institution's system then calculates the applicant's creditworthiness and uses this information to formulate a recommendation regarding the requested loan. This analysis is data-driven, but, according to the interviewees, does not use AI in the form of machine learning, but rather simpler rule-based algorithms. The rules of these algorithms are based on decisions and instructions by the credit institution's management and represent the framework in which the recommendation system operates. The recommendation is presented using a traffic light logic: high creditworthiness is expressed by a "green light", which indicates that the front-office staff should generally approve the loan, while a "red light" indicates that the application should be rejected. Finally, in cases where the system's recommendation is indecisive, a "yellow light" is displayed, prompting the loan officer to contact a risk analyst to review the application individually ("four-eyes" principle). The analyst considers additional criteria and can make an approval decision that is largely independent of the recommendation system. In special circumstances, the bank's staff can deviate from the system's recommendations or even bypass the automated recommendation process, which we will describe in our findings. In addition, the system's recommendation is significantly influenced by the score generated by a credit referencing agency (in Germany, this is usually the Schufa), which is a key indicator used by credit institutions to assess an applicant's creditworthiness. The credit score incorporates a variety of data obtained from different sources (third parties, such as cellular providers) and processed by an internal scoring system. The lending process is a highly complex ecosystem of individual, self-contained process loops, including the creation of a credit score and the development of the scoring system. It is only through the combination of these processes that final credit lending decisions can be made. In this regard, it is not possible to conceptualize the involvement of multiple individuals within a single loop; rather, it is more accurate to perceive these individuals as participating in multiple, intertwined (sub-)loops whose interactions collectively facilitate the completion of an overall process.

It is the inherent responsibility of banks to ensure that their lending decisions are made in accordance with the relevant legislation. Consequently, the legal framework that governs the lending process is of paramount importance for the organization of the processes. Provisions of contract law impose general requirements on the conclusion of contracts, such as Section 491a of the German Civil Code (BGB), which imposes extensive pre-contractual information obligations on lenders [36, 37] or §492 BGB, which requires the written form for consumer credit agreements [36].

However, the requirements resulting from more recent European legislation are of greater interest. The General Data Protection Regulation (GDPR) stipulates the need to implement strict data protection measures with regard to the lawful processing of personal data. For example, justification for data processing must be provided at the outset, and data collection must be clearly delineated and limited to the minimum necessary for processing.

In addition, specific European directives have been established for certain categories of credit agreements. The regulation of consumer credit, which is relevant to our considerations, will undergo significant changes as a result of Directive (EU) 2023/2225 on consumer credit agreements. In addition to extensive requirements regarding the credit assessment process, the Directive includes a comprehensive prohibition of discrimination.

The functionality and legality of banking products are periodically reviewed through internal and external audits. The aim of the audits is to ensure that the products comply with the law. External audits are conducted by auditors, the BaFin or the European Central Bank. Internal audits are performed by the Bank's Internal Audit Department, which conducts ad hoc and thematic audits. These audits examine staffing levels, compliance with regulations and the implementation of new products or processes for possible gaps. If an internal or external audit identifies weaknesses or a need for optimization, potential problems are promptly addressed.

From a methodological point of view, the case of lending is not an easy object of study, as the insight into the decision-making ecosystem is only fragmentarily documented for public access, and interviews can only provide a partial description. The aggregation of distributed knowledge sources into an architecture model based on the ARBITER method [29] was therefore an important step towards an overall understanding of the complex interplay. This also explains why we developed a new methodological design based on the described combination of existing methods. Our approach will be repeated and iterated in subsequent case studies beyond this specific case in at least three different case studies. We believe that the proposed methodological approach will also make a significant contribution on its own, but we plan to reflect on our methodological design only after further iterations, accumulating knowledge from future case studies.

3 FINDINGS

In the following, we present our findings on the different Hilo contributions to the credit lending process, the different roles they encompass in it and the influencing factors that affect the quality of the decision-making process.

3.1 Humans-in-the-loops in the credit lending field

First of all, our research validated the correction of the Hilo concept that has already been stated by other researchers. In the early stages of the project, our approach was based on a traditional, regulatory understanding of the Hilo concept. It was assumed that a specific human actor would be identified and interact with a specific machine-generated decision, primarily in the form of a specific recommendation. It soon became clear that this assumption was premature, and that we could identify many actors influencing the decision process. However, our analysis of the process showed that front-desk advisors, and risk analysts are the actors most clearly acting as humans in the loop as described by Crootof et al. [1]. They argue that in practice, more than one person is usually involved in the Hilo function, and most of the time, it is also more than one loop that needs to be considered for the decision-making loop closest to the recommendation system; so, our analysis focused primarily on them. As referenced above, the Hilo literature refers to specific roles that human actors can fulfill. In our understanding of the loan origination process, several such roles can be attributed to Hilos. In particular, we found that front-desk advisors play an interface role, helping to ensure that the consumer's financial information is entered into the application in the correct manner to provide a sufficient basis for decision-making. This role is critical, as consumer input errors are often cited as a barrier to successful credit applications.

As mentioned, the framework of potential credit decisions for a given individual case is first shaped by decisions made at the management level, which are then implemented in the system by IT staff. These management decisions are influenced by external factors, including economic considerations (e.g., ECB interest rates), legal constraints (e.g., anti-discrimination frameworks or requirements for the implementation of automation systems in the lending process), and social factors (e.g., the impact of skills shortages). Experts in our interviews stressed their understanding that credit lending decisions, despite automation support, are made by humans:

> "Of course we use a rating system to assess creditworthiness. This is a regulatory requirement. But at the end of the day, it's the individual who makes the decision. Of course, the rating result is a factor that flows into the decision. But people make the decisions. Example: Despite a poor rating result, the analyst may have reasons that put the rating result into perspective. Then he approves the loan. Or vice versa: Good rating result, but there are other points against granting the loan." *Expert for credit risk management, translated by authors*

For recommendations to work as expected, it is of utmost importance that the applicant, starting the credit decision process, provides the required financial information accurately and correctly in the designated interface. In particular, cases that deviate from the standard inputs can be challenging to categorize within the predefined data entry fields, as the following statement underlines:

> "I think the big problem is that the consumer doesn't understand it. The masks don't even allow that. So,

> with all this automated application through these websites etc. or apps, there is no way to enter a free text field. [...] And if you now clearly define that we only have the data fields XY and there is information beyond that, but it is not requested or the customer cannot write it anywhere, then their situation is not fully captured. I think that's the problem. The data situation is good, but not good enough." *Internal banking expert, translated by authors.*

Currently, consumers can seek assistance from front-desk staff in individual branches when completing loan applications. The central role of data entry in the credit decision-making process pushes credit applicants to take on a human-in-the-loop role themselves, especially in online applications, which could increase with further automation. If they are the only humans left interacting with the system, their inputs and expertise becomes crucial. The results of the interviews indicated that data entry interfaces are not yet able to adequately capture certain cases, especially those where the relevant financial information cannot be sufficiently translated into input data. In several interviews, front-desk advisors mentioned cases where they initiated an alternative manual process if the system did not allow the accurate data to be entered. One gave the following example:

> "It worked for me. Parents on parental leave. Now that I think about it, I never thought we'd do a home loan like that. But then I had an employee who said: Yes, that's understandable, so just do everything manually. So, as long as there's a little bit of human touch, it's possible to make decisions outside the norm." *Internal banking expert, translated by authors.*

Consequently, the responsibility for decision-making in these cases remains with Hilos. The quality of their decisions is valued here (also by peers), because it does not depend on the presentation of financial information within predetermined dimensions but requires human interpretation and a solution beyond the automated systems paths, as another statement exemplifies:

> "I recently had a similar case. It involved a couple where he accepted a severance offer from his company because he was about to retire. The pension starts in exactly ten months and the severance pay is enough to bridge this transition period without any problems. However, such a case does not fit into any existing rating system. A human must actually assess the situation here, as an AI cannot currently do this. [...] From the machine's point of view, he currently has no income. If such circumstances are not incorporated into the system in a different way, the AI will always come to a negative decision. Of course, this can be overruled, but human judgment remains crucial here."[1] *Business Customer Advisor in a German Bank, translated by authors.*

[1]Another interesting observation we made from statements like this is that even staff strongly interacting with the automated credit recommendation system, referred to it as *AI*, while management staff explained explicitly that the systems are currently rule based algorithms and not based on machine learning. We will address this finding in section 4.2 as an influencing factor.

This human information processing, as it is currently conceptualized, is not yet adequately represented in the roles described by Crootof et al. [1] or others. It also exceeds what Raso [38] calls humans "manipulating the system" as a means to interact as skilled decision-makers, since front desk employees and especially risk analysts have an organizational mandate to decide based on their expertise and overwrite the automated recommendations. Consequently, we propose an additional role, which we call the 'special case handling role' of humans in the loop. It captures the mandate and ability of humans to identify when certain cases fall beyond the automations systems' decision paths or other capabilities and therefore realize a different handling of this specific case.

Another statement exemplifies that human experience combined with this mandate to possibly overwrite or circumvent the automation of credit recommendations, can also lead to negative outcomes for the customers:

> "I often hear that in the credit department. Here's an example: I'm talking about a customer who operates in sector X. From my point of view, everything is fine. Then someone comes along and says: 'Oh, sector X – we've had a default there before. That's why we don't do it. A machine doesn't have this subjective 'flavor'. The machine only looks at figures, data and facts, without allowing itself to be influenced by such soft factors. This can be an advantage, but of course it can also be a disadvantage. As is so often the case, it can go either way." *Business Customer Advisor in a German Bank, translated by authors.*

Scholars such as Lobel, or Kleinberg et al. have argued that human decisions under certain conditions can be biased and result in worse decision-making than full automation [39, 40]. Despite the awareness of this factor, in the credit lending sector, experts often express their recognition for expertise through experience, as this statement illustrates:

> "Experience should not be underestimated. An experienced person who has been doing this for ten years knows very quickly where to look. An inexperienced junior employee, on the other hand, naturally takes longer." *Expert for credit risk management, translated by authors.*

Furthermore, although Crootof et al. [1] assign a corrective and a resilience role to Hilos neither currently fully encompasses a process design in which the machine system escalates a decision to a Hilo in cases where it does not reach a sufficiently reliable assessment. We propose to extend the understanding of the resilience role, to capture this process case we found and, in the following, explain, why this case goes beyond existing descriptions of the corrective and the resilience role:

In the context of credit scoring systems, a yellow light decision is displayed when the system does not reach a clear yes or no recommendation regarding the loan application, prompting the decision to be escalated to a risk analyst to reevaluate the case and make the decision. This substitute function is not yet covered by the corrective roles of Hilos proposed by Crootof et al., which so far in their description include [1], correcting system decision errors (1), adjusting the system's output in specific situations (2), and correcting a system's biases (3). The escalation of a particular application to the risk analyst does not directly correct the system's output or error nor its bias, but introduces a new decision path. Crootof et al. also consider humans in the loop in a resilience role, where human actors make decisions on behalf of the system in the event of malfunction or failure [1]. However, this limited scope does not fully address the aforementioned yellow light escalation either. The credit scoring system escalates to risk analysts in cases of uncertainty without experiencing a malfunction or breakdown. Consequently, this paper proposes to extend the understanding of this resilience role to include instances of *voluntary delegation* of authority to a human-in-the-loop.

Overall, our empirical data show that within the range of consumer credit that we have been able to analyze, the role of human expertise and control is still quite extensive and also highly valued socially, as this statement exemplifies:

> "And then, of course, credit institutions have a certain amount of discretion in their lending decisions, for both real estate and consumer loans. There are credit institutions that also place a certain value on personal decisions. For example, at a recent conference, a representative of a credit institution said that even for consumer loans, if the bank employee has a bad gut feeling and although everything looks good, he does not want to conclude the loan agreement or argues against it, the bank will do the same, i.e., the loan application will be rejected. Although, as I said, this is more the case with real estate loans than with consumer loans". *Financial researcher, translated by authors.*

Even more relevant are the influencing factors on humans and humans in the loop in particular, which we will introduce in the following.

3.2 Influencing Factors

For the analysis of this case study, we distinguish between determinants and influencing factors. We define determinants as variables that directly determine the outcome of a decision (often in an if-then logic with defined weights). Thus, the outcome of the loan application depends on the assessment of the specific data elements entered, such as income and repayment history of previous loans. Influencing factors, on the other hand, are variables that indirectly influence the outcome of a decision in a less clear, more complex way. These variables may affect the entire decision process and may also influence prior decisions made by the same or other internal or external actors. Such factors include, but are not limited to, environmental considerations of a political, economic, social, technological, environmental, or legal nature. In this context, it is important to consider the individuality of the decision-maker, as personal factors—particularly social and psychological—can have a significant influence. This also includes factors that influence the selection and prioritization of determinants. It is important to note that, unlike determinants, these factors do not follow a universal if-then logic. In the following, we focus on the factors that influence the (according to our findings) most relevant people in the loop in the context of credit decisions: front-desk advisors, who primarily embody the interface and corrective roles [1], and risk analysts, who primarily ensure the resilience role [1]. However, we believe that many of these factors are relevant in other contexts and can be partially generalized. The factors we identified can be grouped into three levels: the external level, the actor level, and the technological level of the system.

First, we were able to identify external factors. These are located on the outer edge of the graph in Figure 1. These external factors have a more general impact on an organization and its environment and include societal factors such as the shortage of skilled labor or the perceived need for greater automation of credit scoring. In addition, the overall economic orientation and organizational structure of the credit institution, as well as its assessment of economic incentives and the applicable legal framework, also influence the definition of the framework for possible credit decisions by front-desk advisors and risk analysts.

Second, at the actor level, we were able to identify a total of seven factors that significantly influence the decision-making process of the described Hilos and have a direct impact on the individual case decision. These factors are graphically represented in Figure 1 as social factors surrounding the two Hilos. They include the following:

1. **Interpersonal contact** (front-desk advisors only): Front-desk advisors are responsible for direct interpersonal contact with loan applicants, in which they play a facilitating role in the application process. These interpersonal interactions between consumers and front-desk advisors can help minimize potential errors in the application process and account for exceptional cases that fall outside the automated decision parameters. Our findings suggest that for complex decisions, i.e., those involving multiple decision-relevant factors, prior consumer education, combined with the option of asking human agents for help with the application, is paramount.
2. **Decision latitude** (risk analyst only)**:** The quality of the risk analyst's credit decision depends on the degree of decision latitude. The decision is influenced by the distinction between performing a reassessment independent of prior system recommendations or following automated recommendations. In the context of lending, the scope latitude for decision making by risk analysts appears to be substantial and comparable to a reassessment of financial data. However, it is important to note that this factor does not *per se* guarantee improved decision outcomes. Rather, it is critical that the decision latitude be reflected and consciously designed to optimize the human-in-the-loop function.
3. **Understanding of the credit scoring system:** The decision quality of both Hilos is influenced by an accurate understanding of the capabilities and limitations of the technical system that provides the recommendations on the applicant's creditworthiness. Knowing, for instance, if the system is rule-based or based on a trained machine learning model as well as understanding the implications of the difference (which was often not the case in our field of study) is imperative to judge the system's capabilities. This allows the system's recommendations to be more accurately categorized and, if necessary, challenged.

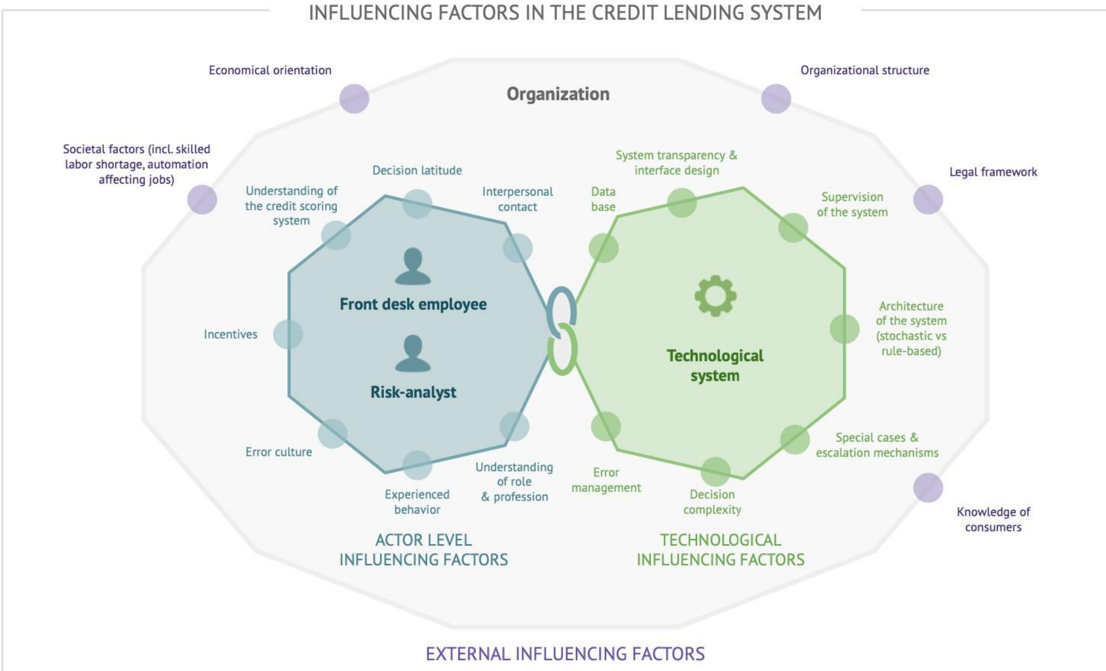

Figure 1: Influencing factors for Hilos in the credit lending system

4. **Incentives from economic orientation and organizational:** The individual decision of the Hilos is also influenced by the internal incentive structures of the credit institution. These can be derived in particular from the organizational structure. For example, a Hilo’s willingness to take risks may depend crucially on the variables on which a promotion or bonus payment depends. If the accuracy of forecasts is more decisive, a more cautious approach may be taken than if the total volume of credit granted to the Hilo is more relevant.
5. **Error culture:** The prevailing error culture in a credit institution can influence the quality of decisions made by frontline staff and risk analysts. The key question is what is considered an error in the decision-making process, such as potential discrimination. This determines which problems can be addressed and escalated, if necessary. It also plays a role whether Hilos have to fear serious consequences for individual wrong decisions, so that they may scrutinize an individual decision longer and may be less willing to deviate from the automated recommendations if this has negative consequences. Finally, how weaknesses in the decision-making process are handled is also relevant: are they discussed openly and improvements are sought collectively, or is internal criticism of the system undesirable?
6. **Experienced behavior:** In our interviews with experts, it became clear that many years of experience, e.g., as a consultant or risk analyst, is viewed positively and valued in credit institutions. According to these descriptions, the experience of risk analysts can accelerate the decision-making process and significantly influence the outcome of a decision, both positively and negatively. Experience provides guidance in the case of similarities. It should be noted, however, that under certain conditions, routine can always lead to errors or even bias due to a lack of reflection in the application of familiar patterns.
7. **Understanding of role and profession:** In addition, how Hilos understand their role and profession can influence how much trust they place in their own assessments (or automated recommendations) and therefore how often they deviate from automated preliminary decisions. At the moment (at least in our impressions of the credit sector), an ambivalent picture emerges regarding the classification of one’s own profession: while a majority of voices tend to emphasize human expertise as central and superior, others emphasize what they see as the superior “more objective” recommendations of the automated system. Our findings suggest that the way in which human loan officers understand their profession is in a state of flux, in part due to increasing automation.

Third, we identified seven influencing factors that are more directly related to the technological level of the socio- technical-system. These are colored green in Figure 1:

1. **Data base:** Credit scoring is based on a wide range of financial data, including both the applicant’s financial history and external credit scores. The quality of the underlying data is therefore critical to the quality of the credit decision. In addition to content-focused data quality dimensions such as the accuracy or timeliness of the information, the selection of data types used is particularly relevant.
2. **System transparency & interface design:** The transparency of the system and its data processing as well as

the interface design are also relevant. In order to be able to adequately process the information about the facts of life, it must be possible to convert this information into processable data points. The quality of the interface design of the credit scoring system is crucial. In particular, the ability to represent complex and unusual life circumstances depends heavily on this design.

3. **Supervision of the system:** System oversight plays a critical role in the quality of system output. Do the human supervisors have the necessary technical understanding and information base to verify the quality of the technical system's recommendations? Is oversight only incidental to daily interactions with the system, or are specific oversight tasks performed only at specific times? Is oversight assigned to specific personnel or performed by staff who interacts with the system in general?
4. **Architecture of the system (stochastic vs. rule-based):** The technical design of the credit scoring system also plays a large role in the effectiveness of human interaction. Surprisingly, some of our interviews revealed that the technical capabilities of these systems are sometimes significantly overestimated, even by experts directly involved in the lending process. This may have a particular impact on the trust that Hilos place in these systems.
5. **Special cases & escalation mechanisms:** Of particular importance for the quality of the system's recommendations, but also for the trust Hilos places in a credit scoring system, is its ability to react to special cases and to escalate the decision to human actors in case of doubt.
6. **Decision complexity:** Automated systems are generally good at processing large amounts of data. However, the system design must also be able to map the underlying information complexity of a situation. Highly complex decision systems are also more difficult for human actors to monitor [41].
7. **Error management:** Finally, how errors in the technical system are handled is an influencing factor. Are incorrect decisions made by the system adequately documented, analyzed, and improved? If incorrect decisions made by the system are not adequately identified, or if such decisions do not lead to continuous improvements in the system, its decision quality will stagnate. This can have a particular impact on the perceived competence of the system and thus on the willingness to follow its recommendations.

These influencing factors are so relevant because they can also be understood as levers for influencing the effectiveness of human actors. Thus, they offer scope for actively shaping and, in the best case, improving the decision-making system in order to respond to existing problems and challenges.

4 DISCUSSION

There is a strong discrepancy between the abstract regulatory understanding of the human-in-the-loop and the messy reality of (semi-)automated decision systems in complex decision architectures. Our study provides clear added value by reality checking the abstract understanding of the Hilo and identifying pain points regarding their effective involvement. On the other hand, our empirical findings can contribute to a more applied understanding of the Hilo, by providing evidence for a broader definition and a more nuanced understanding of Hilo roles as proposed in this work. Policy makers can use our influencing factors in designing meaningful control requirements. While some of our findings are tentative and subject to further testing, we want to discuss them because of their relevance to the broader field of algorithmic human-machine interaction and can have important implications, including for regulatory approaches. We see four key points to improving the quality of Hilo requirements and realities.

The first two recommendations relate to addressing epistemic failures in the design and communication of human-machine interaction in both intra-organizational and extra-organizational perspectives. The latter two points are derived from certain strengths which we identified in our case study.

A key finding amongst the influencing factors is rooted in employees' lack of understanding of the recommendation systems' architecture, its exact capabilities, workings, and limitations, and alongside this, their own role within this architecture also to improve and challenge it. To ensure (partly) automated architectures work smoothly, **(1) the decision architecture must be clear to those involved in making decisions.** Beyond training within credit institutes to provide employees with this knowledge, we suggest that it would be helpful to appoint an actor to oversee the entire process and continuously share knowledge with relevant stakeholders. Such an actor seems essential to enable stakeholders to better reflect and understand their role in the interplay of organizational dynamics and automation, and to identify potential problems or opportunities for improvement. Related to this, but a separate effort, we see greater transparency to customers and their advocates as a helpful improvement, since our data shows that the customers responsibility for the workings of the decision systems architecture is influenced by his entries and thereby his knowledge which data influences the outcomes (and how). Therefore: **(2) The decision architecture must be made clear to those affected by the decisions.** According to banking experts, consumers have insufficient knowledge of the criteria used in credit decisions. This applies both to understanding the general relevance of certain criteria and to identifying which specific criteria influenced their individual case. Many experts stressed the importance of greater transparency in the decision-making criteria and improved financial education for consumers as necessary and beneficial measures. In a very recent judgement (C-203/22) the European Court underlined this necessity as well, when it found that data processors "must describe the procedure and principles actually applied in such a way that the data subject can understand which of his or her personal data have been used in what way in the automated decision-making at issue" (Recital 61).

The case of credit lending also provides interesting learnings from its architectural perspective: The integrated option to escalate cases to humans (signaled by the yellow light) is a valuable feature for this very complex field of decision making. This escalation mechanism is in place for an uncertain recommendation, which leads to the "four-eyes" principle and escalates the decision from the front desk to the risk analysis, which ties in with Crootof's (extended) resilience role [1]. This built-in function within the system to identify cases where deeper human control is needed is a mechanism that decision makers and system designers should be aware of as a best practice in decision scenarios that often produce complex and context-dependent outcomes.

We believe this option should be considered more often in Hilo scenarios: In cases of uncertainty or other situations, where this makes sense, **(3) machines should be able to call on the humans in the loop**. For many situations also beyond the credit business scope, it is critical for human-machine interaction that machines function within certain limitations and are able to fully or gradually escalate decisions to humans or increase automation depending on the situation. The question of whether automated decision systems should be able to independently vary their level of automation within decision processes is critical to the development of safe and reliable technologies. This adaptive and flexible automation allows prioritizing human control and also human accountability at critical moments, especially when the system detects anomalies or encounters an uncertain situation, ensuring resilience figure 1]. Finally, **(4) the ability to provide exceptions to the use of automated systems, especially when dealing with non-standard settings is an important mechanism to consider from a regulatory and design perspective.** These expectations to bypass the automated system due to shortcomings to address case specificities in the field of the credit business are allowed and valued and in our case study occur for example because certain data cannot be included in the analysis. These exceptions played an important role in ensuring equal treatment and solutions to unconventional data inputs representing a living situation of a customer. In a fully automated system without this ability to allow exceptions, these cases would most likely have led to inappropriate decision outcomes.

5 CONCLUSION AND OUTLOOK

In conclusion, the credit business as an important socio-political field of action provided interesting first results on influencing factors for the humans-in-the-loop in practice and allowed us to derive learnings for the normative use of the concept in legislation. In particular, our findings on the handling of exceptional cases beyond automation and the problem of a missing decision architecture overview within the organizations are important contributions to the discussion on human-in-the-loop implementation. This case study shows that successful automation requires not only technical optimization, but also a deeper understanding of the human roles within these systems, and in particular the implicit knowledge that humans hold within organizational structures. Building on Crootof's seminal circumscription of Hilo-roles, we suggest the addition of a special case handling role, and argue for a broader understanding of the existing resilience role. Our research emphasizes the importance of looking at decision-making systems, their technical and non-technical actors, and their interrelationships, and not just at individual decisions. Our study makes clear that at the current level of automation in credit lending, human knowledge and experience are indispensable, especially for exception handling and error limitation. Automation has potential, but requires a deep understanding of socio-technical systems in order to not jeopardize the quality of decision-making. Therefore, for this scenario humans remain the central actors in order to recognize the complexity of individual decisions and to ensure the quality of the decision-making process. This is not to say that people are always the (better) solution. Rather, we want to draw attention to the interaction between humans and machines and emphasize that it is crucial to understand, utilize, and harmonize the roles and capabilities of both human and technical actors in order to implement decisions according to the desired criteria. Processes of human-machine integration could greatly benefit from informed benchmarking against system design alternatives (e.g. fully automated vs. partial human oversight), tailored to relevant success indicators for the respective scenario. Such an approach was beyond the scope of the present study, but we encourage future research to utilize the identified influencing factors in an appropriate construction of benchmarks to respect dimensions beyond credit default rates.

Increased automation and deeper involvement of AI systems in the lending workflow in the future were recurring themes in the qualitative interview data we collected. In our expert interviews, narratives about future automation painted a rather ambiguous picture, with voices ranging from enthusiasm and hope, over acceptance of automation as a given goal, to more worrisome voices. Something that is noteworthy from our data, is that system literacy does not appear to be widespread in the lending industry yet, leaving room for misinterpretation of the system's capabilities, its limitations, and drawbacks. At this point, the implementation of further automation, and in particular AI systems, remains uncertain, and it is still unclear what the consequences will be. One emerging threat we have identified in our research arises from the central role of data entry. If the automation of lending pushes loan applicants to become the central and sole actor as data providers, with less human guidance, support, or intervention from today's existing humans in the loop, this could lead to many unforeseen problems. This constellation could lead consumers themselves to take on a human-in-the-loop role, giving them more responsibility.

From a regulatory perspective, future regulations, such as the national implementations of the (EU) 2023/2225 directive on credit agreements for consumers, will put even more emphasis on human decision-making and the impact of humans on automated decision-making processes. Therefore, it is essential to understand the real-world environments of these decisions, which our paper contributes to.

As the first of four case studies, we believe this research is promising as it shows that the assessment of human-in-the-loop figurations can be highly insightful and informative from a research perspective as well as for policy work. We hope our research encourages similar case studies in other regions and fields of human-in-the-loop scenarios. This would lay the foundation for further comparative research. We also plan to expand our scope and repeat our approach to Hilo case studies with different setups, for example in the field of content moderation or aviation. Looking at different actor constellations and decision-making processes will be helpful to develop a more abstract understanding of the influencing factors and allow us to further evaluate and extend our findings to enrich the debate about the human-in-the-loop with insights from real life cases.

Acknowledgments

This project is funded by the Stiftung Mercator. We thank Lara Kauter and Sadiem Youssef for their support as student assistants on this paper.

References

[1] Rebecca Crootof, Margot E. Kaminski, and W. Nicholson Price. 2023. Humans in the Loop. Vanderbilt Law Review 76 (2), 429–511. http://dx.doi.org/10.2139/ssrn.4066781

[2] Ben Wagner. 2019. Liable, but Not in Control? Ensuring Meaningful Human Agency in Automated Decision-Making Systems. Policy and Internet 11(1): 104-122. https://doi.org/10.1002/poi3.198

[3] Jake Goldenfein. 2024. Lost in the Loop - Who is the "human" of the Human in the Loop? Cambridge University Press 2024. https://ssrn.com/abstract$=$4750634

[4] Ximeng Chen, Xiaohong Wang, and Yanzhang Qu. 2023. Constructing Ethical AI Based on the 'Human-in-the-Loop' System. Systems 11 (11): 548. https://doi.org/10.3390/systems11110548sf

[5] Ross P. Buckley, Dirk A. Zetzsche, Douglas W. Arner, and Brian W. Tang. 2021. Regulating Artificial Intelligence in Finance: Putting the Human in the Loop. The Sydney Law Review, 43 (1): 43–81. https://ssrn.com/abstract$=$3831758

[6] Therese Enarsson, Lena Enqvist, and Markus Naarttijärvi. 2022. Approaching the Human in the Loop – Legal Perspectives on Hybrid Human/Algorithmic Decision-Making in Three Contexts. Information & Communications Technology Law, 31 (1): 123–53. https://doi.org/10.1080/13600834.2021.1958860

[7] Meg L. Jones. 2017. The Right to a Human in the Loop: Political Constructions of Computer Automation and Personhood. Social Studies of Science, 47 (2): 216–39. https://doi.org/10.1177/0306312717699716

[8] Reuben Binns. 2022. Human Judgment in Algorithmic Loops: Individual Justice and Automated decision-making. Regulation & Governance, 16 (1): 197–211. https://doi.org/10.1111/rego.12358

[9] Kiel Brennan-Marquez, Karen Levy, and Daniel Susser. 2019. Strange Loops: Apparent Versus Actual Human Involvement in Automated Decision Making. Berkeley Technology Law Journal, 745–72. https://doi.org/10.15779/Z385X25D2W

[10] Sebastian Tschiatschek, Eugenia Stamboliev, Timothee Schmude, Mark Coeckelbergh, and Laura Koesten. 2024. Challenging the Human-in-the- Loop in Algorithmic Decision-Making. arXiv. https://doi.org/10.48550/ARXIV.2405.10706

[11] Tianyu Zhao, Mojtaba Taherisadr, and Salma Elmalaki. 2023. FAIRO: Fairness-Aware Adaptation in Sequential-Decision Making for Human-in-the- Loop Systems. arXiv. https://doi.org/10.48550/ARXIV.2307.05857

[12] Naomi Aoki. 2021. The Importance of the Assurance That 'Humans Are Still in the Decision Loop' for Public Trust in Artificial Intelligence: Evidence from an Online Experiment. Computers in Human Behavior, 114 (January), 106572. https://doi.org/10.1016/j.chb.2020.106572

[13] Min Kyung Lee. 2018. Understanding Perception of Algorithmic Decisions: Fairness, Trust, and Emotion in Response to Algorithmic Management. Big Data & Society, 5 (1): 2053951718756684. https://doi.org/10.1177/2053951718756684

[14] Daniela Sele, and Marina Chugunova. 2024. Putting a Human in the Loop: Increasing Uptake, but Decreasing Accuracy of Automated Decision- Making. Edited by Jiafu An. PLOS ONE, 19 (2): e0298037. https://doi.org/10.1371/journal.pone.0298037

[15] Chidera W. Amazu, Joseph Mietkiewicz, Ammar N. Abbas, Houda Briwa, Andres Alonso Perez, Gabriele Baldissone, Micaela Demichela, Davide Fissore, Anders L. Madsen, and Maria Chiara Leva. 2024. Experiment Data: Human-in-the-Loop Decision Support in Process Control Rooms. Data in Brief, 53 (April):110170. https://doi.org/10.1016/j.dib.2024.110170

[16] Elke U. Weber, and Eric J. Johnson. 2009. Mindful Judgment and Decision Making. Annual Review of Psychology 60 (1), 53–85. https://doi.org/10.1146/annurev.psych.60.110707.163633

[17] Hans-Rüdiger Pfister, Helmut Jungermann, and Katrin Fischer. 2017. Die Psychologie der Entscheidung: Eine Einführung. Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-53038-2

[18] Bernd Schwandt. 2021. Entscheidungsprozesse und Emotionen. Wiesbaden: Springer Fachmedien Wiesbaden. https://doi.org/10.1007/978-3-658-35936- 2

[19] Giulio Mecacci (Ed.). 2024. Research Handbook on Meaningful Human Control of Artificial Intelligence Systems. Edward Elgar Publishing Limited, 1st ed. 6

[20] Scott Robbins. 2024. The Many Meanings of Meaningful Human Control. AI Ethics 4 (4), 1377–88. https://doi.org/10.1007/s43681-023-00320-6

[21] Filippo Santoni De Sio, and Jeroen Van Den Hoven. 2018. Meaningful Human Control over Autonomous Systems: A Philosophical Account. Frontiers in Robotics and AI 5 (February):15. https://doi.org/10.3389/frobt.2018.00015

[22] Walter Didimo, Luca Grilli, Giuseppe Liotta, and Fabrizio Montecchiani. 2022. Efficient and Trustworthy Decision Making through Human-in-the- Loop Visual Analytics: A Case Study on Tax Risk Assessment. Rivista Italiana Di Informatica e Diritto, 4 (2), 15–21. https://doi.org/10.32091/RIID0092.

[23] Sarah E. Walsh and Karen M. Feigh. 2021. Differentiating 'Human in the Loop' Decision Process. In 2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC), 3129–33. Melbourne, Australia: IEEE. 10.1109/SMC52423.2021.9658802

[24] Ujué Agudo, Karlos G. Liberal, Miren Arrese, and Helena Matute. 2024. The Impact of AI Errors in a Human-in-the-Loop Process. Cognitive Research: Principles and Implications, 9 (1): 1. https://doi.org/10.1186/s41235-023-00529-3

[25] Pawinee Pithayarungsarit, Tobias Rieger, Linda Onnasch, and Eileen Roesler. 2024. The Pop-Out Effect of Rarer Occurring Stimuli Shapes the Effectiveness of AI Explainability. In Proceedings of the Human Factors and Ergonomics Society Annual Meeting, 68 (1): 352–58. https://doi.org/10.1177/10711813241261284

[26] Iyad Rahwan. 2018. Society-in-the-Loop: Programming the Algorithmic Social Contract. Ethics and Information Technology 20 (1): 5–14. https://doi.org/10.1007/s10676-017-9430-8.https://doi.org/10.1109/SMC52423.2021.9658802

[27] European Commission. 2019. Building Trust in Human-Centric Artificial Intelligence. COM(2019) 168 final. Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions, Brussels, 6.

[28] Federal Constitutional Court of Germany. 1973. Soraya (BVerfGE 34, 269). Decision of February 14, 1973. Federal Constitutional Court, Karlsruhe, Germany.

[29] Fabian Burmeister, Christian Kurtz, Josefine Spürkel, Tobias Mast, Ingrid Schirmer, and Tilo Böhmann. 2024. Interdisciplinary Architecture Modeling for Regulating Digital Business Ecosystems. PACIS 2024 Proceedings, 3. https://aisel.aisnet.org/pacis2024/track06_dpe/track06_dpe/3

[30] Tuğberk Çitilci, and Murat Akbalık. 2020. The Importance of PESTEL Analysis for Environmental Scanning Process. In Handbook of Research on Decision-Making Techniques in Financial Marketing, edited by Hasan Dinçer and Serhat Yüksel, IGI Global, 336–57. https://doi.org/10.4018/978-1-7998-2559-3.ch016

[31] Philipp Mayring. 2014. Qualitative Content Analysis: Theoretical Foundation, Basic Procedures and Software Solution, Klagenfurt. https://nbn-resolving.org/urn:nbn:de:0168-ssoar-395173

[32] Herbert J. Rubin, and Irene Rubin. 2012. Qualitative Interviewing: The Art of Hearing Data. 3rd ed. Thousand Oaks, Calif: SAGE

[33] Cornelia Helfferich. 2019. Leitfaden- und Experteninterviews. In Handbuch Methoden der empirischen Sozialforschung, edited by Nina Baur and Jörg Blasius, Wiesbaden: Springer Fachmedien Wiesbaden, 669–86. https://doi.org/10.1007/978-3-658-21308-4_44

[34] Jochen Gläser, and Grit Laudel. 2010. Experteninterviews und qualitative Inhaltsanalyse als Instrumente rekonstruierender Untersuchungen. 4. Auflage. Wiesbaden: VS Verlag

[35] Udo Kuckartz. 2018. Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung. 4th revised edition. Grundlagentexte Methoden. Weinheim: Beltz

[36] Kai-Oliver Knops. In BeckOGK. Online commentary on the German Civil Code (BGB), §§491a, 492 BGB (margins 12-18). Beck-Online

[37] Christoph Andreas Weber. 2023. In Münchener Kommentar zum BGB. Commentary on the German Civil Code (BGB), §491a BGB. Munich: C.H. Beck

[38] Jennifer Raso. 2017. Displacement as Regulation: New Regulatory Technologies and Front-Line Decision-Making in Ontario Works. Canadian Journal of Law & Society 32(1): 75-95. https://doi.org/10.1017/cls.2017.6

[39] Orly Lobel. 2022. The Equality Machine - Harnessing Digital Technology for a Brighter, More Inclusive Future. PublicAffairs. ISBN-13 978-1541774759

[40] Jon Kleinberg, Himabindu Lakkaraju, Jure Leskovec, Jens Ludwig, and Sendhil Mullainathan, 2018. "Human Decisions and Machine Predictions", The Quarterly Journal of Economics, Oxford University Press, vol. 133(1), pages 237-293. DOI 10.3386/w23180

[41] Nathan R. Bailey, and Mark. W. Scerbo. 2007. Automation-Induced Complacency for Monitoring Highly Reliable Systems: The Role of Task Complexity, System Experience, and Operator Trust. Theoretical Issues in Ergonomics Science 8 (4): 321–48. https://doi.org/10.1080/14639220500535301