Human-Machine Hybrid Decision Making with Applications in Auditing

Sara Hooshangi
Virginia Tech
shoosh@vt.edu

Soheil Sibdari
University of Massachusetts Dartmouth
ssibdari@umassd.edu

Abstract

The decision making process in a variety of organizations faces substantial changes, largely as a result of advances in information technology and artificial intelligence (AI). A considerable number of decisions that were traditionally made by humans, are now made by machines. Consequently, many jobs that were held by experts in some fields are now occupied by data scientists who can build AI algorithms. In this paper, we address this change in work environments and suggest an innovative process for hybrid decision making between humans and machines. We focus on the auditing profession, but our method can be used in other human-intensive and critical fields such as healthcare, financial services public sectors, and humanitarian organizations.

1. Introduction

The 2014 effort by Amazon to automatically review job applications sounded exciting, until it was revealed to be sexist. Reuters, the news agency, broke the story that the AI-enabled recruiting machine does not like women and significantly favors male candidates [1]. Shortly after, this platform was terminated and later appeared in a different automatic talent management system. Since then, several studies have investigated this case and listed many reasons behind this pitfall such as inherent biases in the data and past performance review of employees at Amazon. The incentive behind this effort was not cost saving, and Amazon really wanted to hire good candidates without necessarily repeating previous practices. It was not because of a bias in the data, but rather the outcome of a faulty platform design that thwarted the effort [1]. This system had limited human interference in decision making and minimal human provision of system performance. Adequate human participation in the AI-enabled processes could prevent unacceptable recommendations and save the non-measurable cost that this disruption imposed on Amazon.

In the landscape of the future of work, unexpected biases and errors are foreseen, if the human-technology frontier is not well defined. The recent remarkable advances in the field of artificial intelligence (AI) can lead to situations where a machine-led agent is able to make complex decisions in unknown situations. These technologies can enable a firm to automate a significant portion of the mundane tasks that were traditionally performed by humans at a lower cost. Different applications in the form of Robotic Process Automation (RPAs) are now available to automate many intensive tasks. A cost-saving objective can lead an organization to carelessly adopt these technologies, which makes the work vulnerable to unexpected biases and judgmental errors. One aspect of this frontier is the decision-making process, as a considerable number of decisions that were traditionally made by humans can now be made by machines [2].

In this paper, we focus on presenting a process where machine and humans can work together to solve complex problems. Recently, the paradigm of humans and machines combining their complementary strengths to solve problems has gained attention with concepts such as Hybrid Intelligence [3] or Human-in-the-loop AI (HAI) [4]. With Hybrid Intelligence, human learns from AI and benefits by generating new knowledge about a complex system, and in return transfers implicit knowledge from expert opinions to enrich the AI performance. Human-in-the-loop systems use AI to process large datasets while leaving the complex tasks to humans [5].

The focus of our paper is on the design of the initial phased used in the decision making process before the actual human-machine interaction stage. The use of Hybrid Intelligence methods can be a suitable extension of this current work. To put our process into the context of a real work scenario, we have chosen the auditing profession as our use-case, but our findings potentially can be employed in other fields with a human-technology frontier. The auditing profession is a crucial and human-intensive profession that spreads across many fields such as healthcare,

URI: https://hdl.handle.net/10125/79356
978-0-9981331-5-7

public sectors, manufacturing, financial sectors, and government. Prospective auditors are expected to have the ability to adapt these technologies with business applications, create new levels of process automation, and customize current applications to capture a firm's needs and goals. The participation of auditors in the US labor market is significant, and work improvements in this profession will impact a wide variety of workforces. Auditors play important roles in different sectors as general or internal auditors, project control analysts, controllers, and compliance analysts. According to the US Bureau of Statistics [6] the number of accountants and auditors in 2020 was 1.27 million, below its peak of 1.8 million in 2012, with a medial pay of around $74,000 per year. Auditing is a tedious job with tight time constraints, high responsibility, and high cost per error. Offering a high-quality audit is essential for this job, and many organizations are considering AI technologies to lower the cost of offering such services [7-8].

The exposure of the RPA technology has influenced auditing education as well. According to a survey conducted by the Graduate Management Admission Council [9], there is a decline in the demand for graduate-level accounting degrees. This survey consists of 306 business schools and related institutions worldwide with 1,085 graduate management programs. The number of applicants has dropped significantly in recent years. Interestingly, this decline coincided with an increase in the number of applications to Master of Data Analytics programs. This may show a significant change in the future of education and work in the fields of accounting, auditing, and data science.

Although there are some efforts to implement RPAs in the auditing profession, to the best of our knowledge, no research effort has studied its micro-level decision making process and the socioeconomic aspects of this integration. The literature also faces a gap in the design of a partnership process in mutual decision makings in order to achieve an organizational goal. By addressing the research questions in this proposal, we try to highlight these issues and provide guidelines for the new adventure of human partnership with machines.

In this paper, we provide several processes and procedures that could lead to address the following questions.

- Q1: What are the dominant auditing processes in different fields? What crucial decisions are made in these processes, and what is the cost per error for each decision?
- Q2: Can an auditing job be decomposed into smaller and simpler tasks, and a final auditing decision be made by summing up these decisions?
- Q3: How can the tasks (and the decisions) be distributed between the human and the machine in order to improve the organizational goals and reduce bias and error?

The contribution of our paper is in introducing possible methodologies to address the above questions and indicating the advantages and disadvantages of each methodology.

2. Conceptual background

Our project aims to bridge three areas: (1) Work analysis literature focused on defining jobs, decision making, and task decomposition; (2) Machine and human decision making and their biases; and (3) Socioeconomic studies of new technology. In this section, we first briefly review the existing literature related to our research, and then define terminologies that will be used in our methodology. We used a narrative (traditional) literature review method where we researched relevant databases and keywords to highlight significant areas of research in the AI field as related to our research questions.

2.1. Relevant work

Recently, many researchers have studied the unexpected consequences of relying on AI technologies in human lives [10]. Examples of such consequences are the judgmental errors, systematic and random biases, and the heavy use of machines for cost-saving purposes. Bogost [11] studied AI-enabled systems and referred to faulty interactions between humans and machines as the major source of the errors and biases. For example, inappropriate human involvements in the training phase of an AI algorithm induces the human attitude into the learning phase, which potentially becomes the source of an algorithmic bias. A heavy involvement of machines in the decision making process also leads to another type of bias, the automation bias. Automation bias refers to systematic errors that a machine makes while the human offers little or no provision due to limited human involvement in the process design.

Goddard et. al [12] studied the algorithmic errors in commercial recreational systems and listed possible errors in algorithmic decision making that can lead to high global security risks. They reviewed the literature and explained the risks associated with uncritical reliance on algorithms and automated decision making in credit, financial services, housing, and employment. According to Goddard et. al [12], possible biases of automation can be categorized as follows: (1) Artificial agents autonomously learning from human biases and mistakes; (2) Biases encountered when working with policy or social questions and the difficulty of identifying ground truth when facing a new situation. Human criteria for judging correctness are often culturally or socially informed. The partnership takes place when learning algorithms would be optimized over time by imposing some measures of social acceptability and by internal organization norms that are incorporated by humans; (3) Dealing with a fuzzy, rather than a well-defined, set of criteria. Humans can navigate through a fuzzy cultural norm, complex fuzzy relationships such as government laws and rules, and conclude with subjective evaluations. Capturing such important information requires more than data and machine, and humans are needed to interfere.

In recent years, the notions of fairness in AI and fairness-aware algorithms have received attention in the AI research community [13]. Broadly speaking, AI algorithms can be fairness-aware if they employ techniques to reduce bias and discriminatory outcomes [14-15]. This can be accomplished by using pre-processing techniques that would alter the training dataset in such a way to decrease bias in the outcomes, algorithm modification, or new models to increase fairness in classification, or by using post-processing techniques to modify the output to be fair [13].

The effect of industrial robots and computer-aided technologies on labor markets has been initially addressed by Keynes [16]. In this seminal work, he used the term "technological unemployment" referring to the effect of new technologies on wages and labor market. Since then, many studies analyzed this, including Dhar [17], Graetz and Michaels [18], and Webb [19]. They mainly focused on the variation in robot usage in different sectors of different countries and concluded that although industrial robot usage lowers the employment of low-skill workers, it increases their productivity and wages. More recently, Acemoglu and Restrepo [20-21] addressed this problem and developed a mathematical model in which industrial robots competed against humans in the production of different tasks. They estimated the impact of robots on human employment and wages using a cross sectional analysis over multiple sectors and countries. According to their analysis, adding one more robot per one thousand workers reduces both employment and wages by less than 0.5 percentage point.

In addition to industrial robots, the impact of cognitive automation on human employment is also studied. Manyika et al. [22] used multidisciplinary research in economics and management to develop a micro-to-macro methodology to examine the microeconomic trends in the industry on a broad range of macroeconomic forces and business strategies. They considered six different themes including productivity and growth, natural resources, labor markets, the evolution of global financial markets, the economic impact of technology and innovation, and urbanization. Manyika et al. [22] first addressed that the potential to implement technical automation is high, but unlike industrial automation, cognitive automation can improve the labor market and wage levels. Webb [19] developed a new method to predict the impact of any technology on the labor market using a text-mining method. He introduced a measure of exposure of different tasks to automation using the overlap between the text of job task descriptions and the text of patents. Manyika et al. [22] applied this measure to historical cases such as software and industrial robots in addition to AI and cognitive technologies and concluded that AI technologies can reduce the wage gap and inequality but will not affect wages among top-income employees. In the following section, we introduce the work environment and highlight the impact of machine and automation in decision making. We depict the future of work, technology, and worker after the machine integration in decision making.

2.2. Terminology

To conduct research in human-machine interaction in auditing, we first need to define the new work context and define the future of workers, machines, and the work in the auditing profession. In this section, we will examine each of these domains and explore in more details about the required skills needed to conduct research in this area.

2.2.1. Work context. In this paper, the work is an auditing job, which can be defined as the examination of certain contents belonging to an object in order to present a fair view of a concern. Here, the content can be presented in the form of documentation, books, online activity, or financial statements; the object can be a firm or a person; and the concern refers to use case items and a collection of documented terms and conditions. All objects operate indefinitely in accordance with the concern, until evidence to the contrary is provided. In this context, the worker, an auditor, can verify the accuracy of the subject matter with no technology. However, according to Penn [23] since such work is categorized as predictable and with a relatively high cost per error tasks, automated technology can improve the speed and quality of the work.

Cognitive Work Analysis (CWA) is a well-established approach to model complex sociotechnical work systems. This approach focuses on building a model of how work could proceed in a certain work environment given different constraints. Using this approach, a work system is described, functional properties are listed, the roles of different actors are defined, and the impact of their cognitive skills and strategies is studied [24]. Within this framework, a multidisciplinary approach has emerged to improve the quality of decision making, namely Task Decomposition. Using this approach, a complex task is decomposed into smaller and simpler tasks (i.e., functions) that, later, are integrated to make the final decision [25]. The task aggregation is an important part of this method and refers to a judgmental step to make an overall decision using the sub-decisions that were made for each function. This aggregation needs to be reproducible according to a pre-defined procedure [26].

Our purposed service is delivered using an Intelligent Cognitive Assistant (ICA), which in participation with the worker can better perform an auditing job. Using the ICA, the examination of contents would be divided into mutually exclusive functions that performs similar to the daily activities of an auditor. The size and importance of functions differ from each other and can change over time based on new information, or a change in the organization's goal. In each function, a set of informative items such as numeric measures, statements, texts, or pictures is available in order to learn about the subject matter. The items within each function also differ in their importance and information levels, which are determined by the ICA. According to the function's importance, the ICA reviews a series of items and, after employing machine learning and statistical methods, recommends a few pre-defined decision items. The auditor reviews the items, observes the suggested decision, and selects one according to the use case items. A single-item decision does not require the human interaction and is automatically made by the machine.

Figure 1 depicts an example of our human-ICA partnership. The examination of a function is an auditor goal with pre-defined use case items that is given to both the auditor and the ICA. A set of external and internal data and information will be provided to the ICA and a collection of AI and machine learning technologies will be used within the ICA in order to determine a set of decisions to be provided to the auditor. A single item set means a machine-made decision. The auditor can either select a decision or use the ICA for another revision of the function. If a function is reviewed and successfully decided, the auditor moves to the next function; otherwise, the auditor returns to review more items and redo the functional decision with a new decision set. This process continues until all functions are examined, after which the auditing process is completed and ready for the final report. If successful, then a conclusion about the subject matter is reached, and a report is submitted. Otherwise, the entire process will be repeated.

A value-added outcome of our system is a powerful learning of auditing activities by the ICA. Auditors take different pathways to examine contents, which enables the ICA to recognize the most effective chain of contents to review over time, given the work importance and the auditing goal. This effective pathway is the auditing system proprietary material and could be transferred to other institutions to make an efficient and high-quality examination.

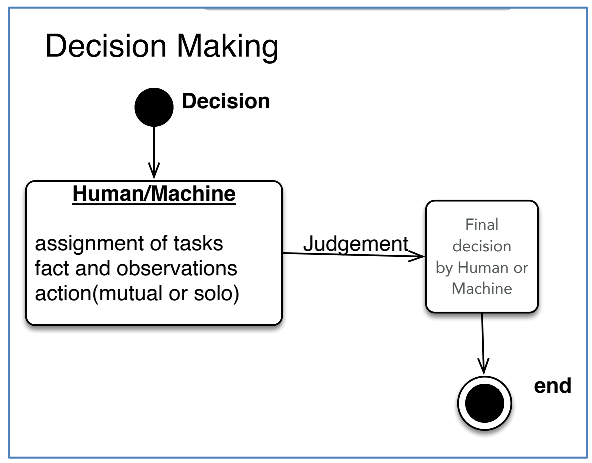

Figure 1: Human-machine decision making

2.2.2. Future technology. Automation creates opportunities for humans to make in-depth judgements and insightful decisions that are more valuable for work and decision making [27]. On the other hand, innovative ideas and decision making provide an opportunity for technology to learn from human values and judgment and incorporate those values into the machine learning algorithms. The future of technology can utilize such hybrid decision making in automation, where AI technology can be augmented to improve available systems. Our hybrid model can create a balance between human and technology roles in the future of work. Another technology that will benefit from our research endeavors is the process design that enterprises can use when implementing AI technology. The hybrid approach in a support system will also impact the design of the user interface in the decision support systems. A flexible enterprise management system that enables an organization to define or redefine the extent of human interaction with an automated technology, can expand the knowledge generation for the future software development industry.

2.2.3. Future workers. The hybrid model of decision making extends the value of auditing jobs into two coherent, though seemingly opposite, directions: less involvement of workers in the mundane tasks, while more participation of human values in the process of auditing. Taking the repetitive and tedious auditing tasks off the workers' to-do-list frees them to think more broadly and focus on other human-based tasks. In the solely human-run auditing jobs, the auditors take up the task at the risk of their own health, as the work involves intensive review of repetitive documents to catch relatively low fraud activities, with relatively high cost per error. This task might have negative emotional and physical health outcomes such as fatigue, isolation, and stress. Unlike fully automated systems, where the human feels less responsible and points the finger at the machine at the time of an error, hybrid decision making keeps the workers in the loop and engages them in a responsible and productive way to participate in the final product, which will be a successful and fair auditing task. The right balance of human and machine involvement in the decision making process transforms the auditing job into a more desirable, responsible, healthy, and engaged job.

2.2.4 Future work. The hybrid model of decision making can encourage firms and organizations to expand their auditing activities, or even try to increase their services or product quality. The hybrid model will remove the burden of high cost associated with hiring human auditors, and the potential resistance of relying solely on an automated machine to perform such tasks. In addition to the auditing-intense industries, such as the financial industry, using this hybrid model in other sectors such as healthcare, publishing, and marketing can introduce auditing jobs into their quality control processes.

International supply chain is another area of work that can be modified using our proposed approach. In today's economy, where a significant number of business activities and supply chains are involved with international trades, the machine alone, or the human alone cannot offer a comprehensive and fair auditing task, since a single trading job can involve multiple countries with different compliance, trade rules and regulations, and cultural factors that a single person or machine is unable to capture and analyze. The hybrid model of decision making enables a firm to decompose the tasks into regional activities, run the auditing jobs locally, and aggregate the outcomes to produce a comprehensive auditing result.

2.2.5. Task decomposition. Task decomposition is not necessarily beneficial. Connolly and Dean [28] and Belli et al. [29] have provided examples of failed task decompositions leading to decision biases. Henrion et al. [30] linked the success of a task decomposition to the way the tasks are decomposed and the way the functions are aggregated. Many other researchers have also studied the benefits and risks of task decomposition [29,31] and their effectiveness factors are task type, the distribution of randomness, and a suitable aggregation algorithm. Lee and Siemsen [31] implemented the task decomposition method in a well-known operational management problem, the newsvendor model, to measure the improvement of the decision when the task is decomposed. In the newsvendor problem, the main decision is how much to order at the beginning of each day when the demand in unknown. They measured the performance using the difference between the order and actual demand and concluded that an ordering system based on a decomposed task performs better than holistic decision making. Rather than using the automated orders, they used the suggested order level of their AI technology as an input for a human, who places the final orders. Task decomposition and decision support systems appear to be complementary methods to improve decision performance in different frameworks.

Task decomposition comes with a variety of advantages and disadvantages compared with other methodologies. Through task decomposition, a big picture about root tasks can be created. This categorization of the task helps to design (or re-design) the decision making hierarchical process. A potential disadvantage of task decomposition is when tasks are placed in incorrect subtasks, resulting in over-estimating or under-estimating of certain tasks which in turn can lower the accuracy of the entire task.

3. Methods, measures, and metrics

In this section, we explain our methodology, and the measures and metrics that are mainly based on task decomposition. We first explain the initial stages of the proposed research including possible interview questions, survey analysis and qualitative methods in order to learn about tasks and subtasks of the work (the auditing job in this research). The qualitative analysis helps to learn about the details of the work and the process design in addition to the importance and sensitivity of subtasks. Later, we explain the quantitative methods in addition to required data and statistical analysis.

3.1. Research preparation and methods

In the initial phase of the work process design, a qualitative survey study needs to be conducted with auditors who work in relevant sectors. This process includes semi-structured interviews to learn about the effect of AI and automation on the day-to-day work of the auditing job [32-33], in addition to learn about variation of the work each auditor faces in the process.

These qualitative studies will help to learn about the first research question, Q1, specifically once a unified set of auditing tasks are identified by the qualitative research study. The outcome of these studies leads to better understanding of auditing tasks from the insights of stakeholders in the field who can share their opinion during open-ended interviews. A series of such questions can be used to learn about the current ecosystem and the future trends in the auditing industry. The survey questions will be guided by three main areas of exploration listed below.

- Current process: What are the dominant auditing processes in different fields? What crucial decisions are made in these processes, and what is the cost per error for each decision?
- Current technology: Is there a significant change in the auditing profession due to AI technologies? What percentage of auditing jobs are replaced by data scientists or computer scientists?
- Error and bias: Do human-based or technology-based decision making on its own impose any judgmental bias or error?

Another qualitative research aspect of this research will measure the auditors' willingness to learn new technologies. This can be done by examining contextual factors that could have an impact on auditors' success and their desire to advance their knowledge in the technology fields. This includes factors such as support provided by the supervisors and the associations' recognition of efforts by peers; desire for rewards associated with success (both emotional and monetary value); level of autonomy for creativity and implementation of new ideas. Table 1 lists possible questions that can be used in interviews with auditors. These questions are merely a sample set and can later be used to construct a survey instrument using survey design techniques.

Table 1. Sample questions for interviews

Tasks
1- Describe the steps you take when auditing.
2- Do you perform your task in a sequence?
3- What is the most error-prone task?
4- What is the most innovative task?
5- How much do you rely on automation and batch processing?
Attitude towards automation
6- What is your experience in using automated system to perform mundane tasks?
7- Do you think human errors be reduced by the use of automation?
8- Do you think human supervision is necessary for all steps in auditing?
9- What is a better model? Complete automation of some tasks, hybrid approach, or entirely human tasks?
Ethical concerns
10- Do you have any ethical concerns about use of automation in auditing?
11- Do you think computers can be unbiased when used in auditing tasks?
Adaptability to new technology
12- How important do you think adapting to new technology is to your professional growth?
13- Have you taken steps towards learning new technologies? If so, which resources do you use?
14- Do you believe AI and automation will be an integral part of auditing profession in the future?

A series of conversations can also be included with different stakeholders (auditors, managers, and university faculty and staff) by conducting focus groups. Themes identified in focus groups will form the foundation of subsequent survey and interview efforts. The topics under discussion may include: 1) the effect of rapid changes in the AI and automation as it relates to the design of effective auditing procedures; 2) potential risks of technology misuse arising from vendor-specific, application-specific, or program-specific circumstances; 3) the need for the reskilling of the auditors to stay current with the technology; and 4) the assessment of the possible need for integration of specialized technologists and other subject matter experts into the auditing industry.

Different mechanisms can be used to recruit participants. Other than the regular call for participation, collaboration with relevant institutes such as the Center for Audit Quality (CAQ) helps to get access to auditing firm personnel in order to participate in the research. The CAQ has a yearly proposal deadline for a program called Access to Audit Personnel program. Upon acceptance, the researcher will have access to CAQ staff who will serve as liaisons between the researchers and the audit firms to facilitate access to the necessary study participants from the firms. The auditors may be contacted for in-person, online, or phone interviews.

3.2. Task decomposition method and metrics

Decision making in complex and highly variant environments is bound to be affected by biases, random judgement errors, coarse decision makings (because of overfitting data), and misidentifying problems. Task decomposition is a well-established method to address these shortcomings and improve the quality of decision making under uncertainty [25]. The benefit of task decomposition is associated with simplifying the problem into separate components, using relevant information about each component one at a time, making judgement about individual components, and finally using an aggregation method to map the judgements into a single decision [26]. This method can also address our research questions namely Q1 and Q2 discussed in Section 1.

The emergence of the information technology field, and the computational capability of AI algorithms have facilitated the implementation of the task decomposition methods. In repetitive and tedious computational work such as auditing, decomposing tasks can help the auditor to focus on tasks requiring human judgment and experience. It also provides guidelines for the human-machine partnership. Here, we introduce a simple example to understand the tentative research studies described below.

Example: Consider an auditing job involved with the risk assessment of a financial statement. The auditor tries to provide a basis for the assessment of the risks by reviewing the documentations and guideline materials. An unethical or fraudulent activity can be detected by comparing the documents with the list of terms and conditions. A machine can also use the historical data to determine specific features and activity leading to a higher risk case. To examine the auditing performance, we use two measures: judgement bias that measures optimistic or pessimistic decision making, and random judgment error that refers to systematic or deterministic errors. These two measures are reproducible, and can be discovered, measured, and corrected over time. The following studies can be conducted based on the above example.

Study I—Judgment bias of the ICA when there is no human intervention

Our main objective in Study I is to examine whether an automated auditing decision is less or more biased than a direct auditing decision with no automated technology. Bias in this example means that the ICA is prone to detect the high-risk activity of certain groups more than other groups. To determine this, we conduct an experiment and compare the performance of a benchmark group (i.e., a human auditor), on average, with the ICA in detecting fraudulent activities within different groups. A judgement bias occurs when the ICA has a higher tendency toward specific actions, or specific groups, which leads to more alerting calls compared to the benchmark. We will also consider the cost per error in determining the bias. The magnitude of Type I error (raising the flag when there is no fraudulent activity) and Type II error (not raising the flag when there is a fraudulent activity) influence the bias level.

Note that cost per error can be found according to a specific application and case study. In auditing, Type I error (i.e., reviewing one additional case) can be calculated using the cost of evaluating one additional case including labor cost (for each case), in addition to possible inconvenience cost that the auditing process incurs including review tardiness. Type II error is potentially more costly because of the cost of approving a fraudulent case. Calculating this cost requires expert opinion in the field. In summary, we hypothesis the following:

Hypothesis 1 (H1). The ICA has a higher tendency of catching risky activities of certain groups compared to the benchmark for an auditing job described in Example 1. This tendency varies according to the cost of Type I and Type II errors.

Study II—Random judgement error of a human

In Study II, we address random judgement errors because human judgement is inherently stochastic [34] In this study, we examine the variation of auditing performance for the same subject (i.e., an auditor), working on multiple jobs (i.e., audits) called within-subject variation compared with the benchmark. The benchmark is when multiple auditors perform audits for different cases. This error occurs in higher rate when the underlying uncertainty of an auditing job is higher (the case requires a greater human judgment). Our Hypothesis is as follows:

Hypothesis 2 (H2). Random judgement error occurs when a single auditor makes multiple auditing jobs. The intensity of this error depends on the uncertainty of the underlying auditing task.

Study III—Task-decomposed auditing with ICA as decision support system (DSS)

In the previous two studies, we examined the performances of the auditor or the ICA separately and addressed different judgement errors or biases. In this study, we measure the auditing performance when the auditing task is decomposed and the human auditor has access to an AI-enabled decision support system, a semi-automated system that allows for human interference at certain points in the process. This design also helps to measure the change in the performance of uninterrupted automated systems (Study I) and the unaided human decision making (Study II) when a hybrid system is in place. For example, the human decision in Study III still makes random judgment errors, but to a lesser extent than Study II scenario. We present the following hypothesis:

Hypothesis 3 (H3). Decomposing the auditing task into separate functions, with human interference at certain points of the decision making, will lead to an improved overall performance.

4. Implementation

To decompose an auditing job into sub-tasks, we use two different literatures in computer science: (1) Work analysis literature, and (2) Parallel computing literature. The work analysis literature uses qualitative and trial and error methods to break down a task into a series of sub-tasks and then convert each sub-task into a single decision or measurable behavior or action. Coffey and Herholz [35] provide a comprehensive task decomposition procedure to examine human brain's capabilities and potentials. In the computer science field, to solve complex problems, parallel computing methods are used, many of which require task decomposition steps. A complex task is decomposed into sub-tasks for concurrent execution, and later they are combined to provide a final solution. Some commonly used decomposition techniques are explained by Grama et al. [36] including recursive decomposition, data decomposition, exploratory decomposition, and speculative decomposition. We will employ methods described in both literatures to provide a suitable procedure for task decomposition in the auditing field. Below we describe the different phases of implementation.

4.1. Phase 1—Qualitative research

Understanding the current decision making process in the auditing industry can be done by surveying auditors and conducting focus groups of industry managers, faculty, and students in the business and data science fields as described in Section 3.1. This involves a series of in-person or online meetings with auditors to conduct semi-structured interviews, which becomes a foundation to build our hybrid model in the full proposal. Table 2 lists the characteristics of auditors best suited for these interviews.

Table 2. Characteristics of interviewees.

Years of experience in the auditing field	2-20 years
Field of work	Financial sector, insurance, healthcare
Years at current position	At least 2 years

4.2. Phase 2—Data collection and analysis

Major findings of this research rely on data collection and analysis. Data can be collected using the qualitative research methods described in Phase 1 and Section 3.1. External sources such as public and private datasets can help to conduct initial investigations of the socioeconomic research. Comprehensive data collection tools such as questionnaires, surveys, and interviews (based on qualitative methods) must be made. During the development of the data collection, the team needs to work closely with key stakeholders to ensure that all instruments will yield reliable and accurate data. Prior to full-scale implementation, the team will pilot-test data collection instruments and conduct preliminary analyses on collected data to verify the reliability of instruments and validity of the findings. The research team may revise instruments based on the pilot results. Depending on the results, such revisions may include modifying the order of the questions, re-translating phrasing of the questions, and/or removing or adding questions.

A variety of secondary existing data sources including labor market data, job postings, and educational related information can be used to complement such work. Using these data sources, the trends in the corresponding labor market can be examined to address possible shifts in auditing and data science job markets. Changes in graduation rate, employment level, and job availability in each of these fields using different tools such as text mining of online job postings can also be performed. These documents will supplement the data collection activities in the field, while also providing references for before-after comparisons.

4.3. Phase 3—Research team recruitment and required skills

A multidisciplinary team is required to further pursue this research and prepare a complete research study. This research spans across many fields in academia and industry, and as such we propose that for such a research agenda to be successful a convergence of researchers from economics, system design, work analysis, and computer science is needed. Creating a framework for human-machine interaction builds on five research skills as follows.

S1. **Qualitative methods** are required to design and conduct the interviews in the planning phase, and the experiments.
S2. **Machine learning and data analysis** methodologies are required to analyze big datasets and provide valuable information in the examination process of an auditing job.
S3. **Quantitative method and system design** expertise are required to design a decision making model for each auditing task.
S4. **Socioeconomic research** is required to determine short- and long-term impacts of this interaction on the labor market and other socioeconomic measures.
S5. **System development** skills are required to implement a technology-based platform.

5. Discussion and contributions

In this paper, we proposed a framework for examining the introduction of AI technology into organizational decision making and associated consequences. Our paper provided: 1) An innovative methodology for the design of decision making process based on decomposing a task (i.e., an auditing job) into mutually exclusive functions with an emphasis on the decision making partnership between human and the machine; 2) A novel procedure to define and assign decision making duties based on the organization objectives and limitations; and 3) A mechanism to provide solutions to the unexpected consequences of technology exposure, such as judgmental bias, lack of accountability, and long-term consequences of replacing experts in the field by data scientists who have less expertise in the subject matter.

Such an approach and research have socioeconomics impacts on the work environment. For example, fraudulent and unethical activities, especially in the financial sector, create substantial social and economic distress that can be prevented if humans participation happens at the right decision making level and the right time. In addition, possible biases towards socioeconomically disadvantaged groups can be addressed and removed by our human-technology partnership solution.

The outcome of this paper has educational values. The new process design and the technology enhancement will guide the computer science field in the development of new AI-enabled products while lowering risks and increasing efficiency. This technology can also be used in other domains such as healthcare, retail management, market research analysis, and more. The outcome of this research can be integrated into the computer science and behavioral science curriculums to better prepare the next generation of workers.

Due to its seminal concept, this paper can be expanded in many directions including its methodology, application, and approaches. Conducting an experimental analysis to learn about the details of auditing task and implementing task decomposition can provide a case study for human-machine hybrid decision making. A proof-of-concept using simulation analysis can highlight the benefit of hybrid decision making and provide measures for comparing different process designs. In another extension of this paper, one can utilize the *Hybrid Intelligence* [5] concept and improve the performance of the process over time using mutual learning of human and machine and their respective interactions. Another extension of this paper can go beyond our decision making application and address more complex task such as problem solving.

6. References

[1] Lauret J., “Amazon’s sexist AI recruiting tool: How did it go so wrong?”, Medium, April 2019, accessed June 15, 2021, https://becominghuman.ai/amazons-sexist-ai-recruiting-tool-how-did-it-go-so-wrong-e3d14816d98e.

[2] Chen, Y. and Zhou, Y., “Machine learning based decision making for time varying systems: Parameter estimation and performance optimization”, Knowledge-Based Systems, 2020, 190, p.105479.

[3] Dellermann, D., Ebel, P., Söllner, M., and Leimeister, J. M. “Hybrid intelligence” Business & Information Systems Engineering, 2019, 61(5), pp. 637-643.

[4] Monarch, R. M. “Human-in-the-Loop Machine Learning: Active learning and annotation for human-centered AI”, Simon and Schuster, 2021.

[5] Demartini, G., Mizzaro, S., and Spina, D. “Human-in-the-loop Artificial Intelligence for Fighting Online Misinformation: Challenges and Opportunities” The Bulletin of the Technical Committee on Data Engineering ,2020, 43(3).

[6] OES (Occupational Employment Statistics), U.S. Bureau of Labor Statistics, May 2020, accessed June 15, 2021, https://www.bls.gov/oes/current/oes132011.htm.

[7] Hanson, J.D., “A call to action for future auditors”, The CPA Journal, 2014, 84(8), p.6.

[8] Ferguson, L.H., “The Importance of Planning and Time Management in Audit Quality”, The 2016 International Institute on Audit Regulation, accessed June 15, 2021, https://pcaobus.org/News/Speech/Pages/Ferguson-audit-planning-Institute-12-13-15.aspx.

[9] GMAC (Graduate Management Admission Council), “Application Trends Survey Report 2020”, accessed June 15, 2021, https://www.gmac.com/market-intelligence-and-research/research-library/admissions-and-application-trends/2020-application-trends-survey-report.

[10] Osoba, O.A. and Welser IV, W., “An intelligence in our image: The risks of bias and errors in artificial intelligence”, Rand Corporation, 2017.

[11] Bogost, I., “How to talk about videogames”, Vol. 47, U of Minnesota Press., 2015.

[12] Goddard, K., Roudsari, A. and Wyatt, J.C., “Automation bias: a systematic review of frequency, effect mediators, and mitigators”, Journal of the American Medical Informatics Association, 2012, 19(1), pp.121-127.

[13] Friedler, S. A., Scheidegger, C., Venkatasubramanian, S., Choudhary, S., Hamilton, E. P. and Roth, D. “A comparative study of fairness-enhancing interventions in machine learning." In Proceedings of the conference on fairness, accountability, and transparency, 2019, pp.329-338.

[14] Romei, A., and Salvatore, R. "A multidisciplinary survey on discrimination analysis." The Knowledge Engineering Review, 2014, 29(5), pp.582-638.

[15] Zliobaite, I. Measuring discrimination in algorithmic decision making. Data Mining and Knowledge Discovery 2017, 31(4), pp.1060-1089.

[16] Keynes, J. M., “Economic possibilities for our grandchildren”, In Essays in persuasion, 1930, Palgrave Macmillan.

[17] Dhar, V., “When to trust robots with decisions, and when not to”, Vol. 17, Harvard Business Review, 2016.

[18] Graetz, G. and Michaels, G., “Robots at work. Review of Economics and Statistics”, 2018, 100 (5), pp.753-768.

[19] Webb, M., “The impact of artificial intelligence on the labor market”, 2019, accessed June 15, 2021, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3482150.

[20] Acemoglu, D. and Restrepo, P.,” The race between man and machine: Implications of technology for growth, factor shares, and employment.”, American Economic Review, 2018, 108(6), pp.1488-1542.

[21] Acemoglu, D. and Restrepo, P., “Robots and jobs: Evidence from US labor markets”, NBER Working Paper Series, 2017, accessed June 15, 2021, http://www.nber.org/papers/w23285.

[22] Manyika, J., “A future that works: AI, automation, employment, and productivity.: McKinsey Global Institute Research, 2017.

[23] Penn, S., “The six-step audit process”, Small Business-Chron.com, 2019, accessed June 15, 2021, https://smallbusiness.chron.com/sixstep-audit-process-17816.html

[24] Rasmussen, J., Pejtersen, A.M. and Goodstein, L.P., Cognitive systems engineering,Wiley, 1994.

[25] Raiffa, H., “Decision analysis: Introductory lectures on choices under uncertainty”, Addison-Wesley, 1986.

[26] Einhorn, H.J., “Expert measurement and mechanical combination”, Organizational Behavior and Human Performance, 1972, 7(1), pp.86-106.

[27] Zappa J., “Human judgment, automation, and the future of ad tech”, StreetFlight, 2019, accessed June 15, 2021, https://streetfightmag.com/2019/03/12/human-judgment-automation-and-the-future-of-ad-tech/#.XmGsuet7m9Y.

[28] Connolly, T. and Dean, D., “Decomposed versus holistic estimates of effort required for software writing tasks”, Management Science, 1997, 43(7), pp.1029-1045.

[29] Belli, R.F., Schwarz, N., Singer, E. and Talarico, J., “Decomposition can harm the accuracy of behavioral frequency reports”, Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 2000, 14(4), pp.295-308.

[30] Henrion, M., Fischer, G.W. and Mullin, T., “Divide and conquer? Effects of decomposition on the accuracy and calibration of subjective probability distributions”, Organizational Behavior and Human Decision Processes, 1993, 55(2), pp.207-227.

[31] Lee, Y.S. and Siemsen, E., “Task decomposition and newsvendor decision making”, 2017, Management Science, 63(10), pp.3226-3245.

[32] Issa, H., Sun, T. and Vasarhelyi, M.A., “Research ideas for artificial intelligence in auditing: The formalization of audit and workforce supplementation”, Journal of Emerging Technologies in Accounting, 2016, 13(2), pp.1-20.

[33] Zhang, C., “Intelligent process automation in audit”, Journal of Emerging Technologies in Accounting, 2019, 16(2), pp.69-88.

[34] Budescu, D. V., Wallsten, T. S., and Au, W. T.,” On the importance of random error in the study of probability judgment. Part II: Applying the stochastic judgment model to detect systematic trends”, Journal of Behavioral Decision Making, 1997, 10(3), pp.173-188.

[35] Coffey E. B. J and S. C. Herholz, "Task decomposition: a framework for comparing diverse training models in human brain plasticity studies", Frontiers in Human Neuroscience, 2013, 7, p.640.

[36] Grama A. and G. Karypis and V. Kumar and A. Gupta, “Introduction to Parallel Computing”, Pearson Education, 2003.