

# Artificial Intelligence Project Success Factors: Moral Decision-Making with Algorithms

Gloria J. Miller  
*Managing Consultant*  
*maxmetrics*

Heidelberg, Germany  
<https://orcid.org/0000-0003-2603-0980>

**Abstract**—The algorithms implemented through artificial intelligence (AI) and big data projects are used in life-and-death situations. While research exists to address varying aspects of moral decision-making with algorithms, the definition of project success is not readily available. Nevertheless, researchers place the burden of responsibility for ethical decisions from AI systems on the system developers. Using a systematic literature review, this research identified 71 AI project success factors in 14 groups related to moral decision-making with algorithms. It contributes to project management literature, specifically for AI projects. Project managers and sponsors can use the results during project planning and execution.

**Index Terms**—artificial intelligence, algorithms, moral decision making, critical success factors, project management

## I INTRODUCTION

ALGORITHMIC decision-making is replacing or augmenting human decision-making across many industries and functions [1, 2]. The decisions range from trivial to life and death. For example, marketing decisions are insignificant compared to legal decisions that may result in incarceration for defendants or the loss of life for health decisions affecting patients. An “algorithm is a defined, repeatable process and outcome based on data, processes, and assumptions” [3]. Algorithms are usually the result of artificial intelligence (AI) or big data projects. It is anticipated that AI will significantly impact society, generating productivity and efficiency gains and changing the way of work [4]. Given the considerable impact on individuals, society, and the environment, understanding the success factors in AI projects is critical.

Sponsoring organizations invest in AI projects expecting them to deliver measurable, meaningful benefits such as revenue or productivity gains [5]. The benefits of AI projects are usually realized long after the projects are completed and the algorithms are put into use. However, the on-time and cost limits of the task or the goal orientation of projects create the risk that the interests of significant stakeholders may not be considered. Thus, the short-term project objectives compared to the long-term social and environmental consequences raise essential questions about the definition of project success.

The decisions or results of the algorithm are what affect the individual and society. The development of large-scale AI models is what affects the environment. Thus, the definition of project success from the public’s perspective should be based on quality, morality, or fairness. The technology view of moral decision-making with AI does not consider non-technical stakeholders, e.g., operators and the public [6]. Manders-Huits [7] explains that the notion of consequences and the level of autonomy of action are preconditions or considerations for moral responsibility, arguing that the burden of responsibility for moral decisions is on the system designers’ shoulders. Martin [8] makes a similar argument stating, “Developers are those most capable of enacting change in the design and are sometimes the only individuals in a position to change the algorithm.” Thus, while research exists to address varying aspects of moral decision-making with algorithms, the definition of project success is not readily available.

The project management literature clarifies that many project stakeholders measure success at different periods and do not share views on the success [9–11]. While the literature acknowledges the importance of client consultation and client acceptances as critical success factors, the public is not foreseen in an active project role. Furthermore, morality is not considered an independent project objective. However, [12] argues that managers should serve legitimate stakeholders’ legal and moral interests.

This research uses a systematic review of the literature to answer a novel question regarding the success factors in AI projects: *what are the project success factors for moral decision-making with algorithms?* It closes the gap on a lack of literature that translates the AI ethical principles into practice [13]. Furthermore, the management of AI projects is hardly covered in the project management literature. This research contributes to the literature on success factors, specifically for AI projects. The paper is structured to provide a literature review, description of the methodology, findings, discussion of the research questions, and conclusions.

## II. LITERATURE REVIEW

### A. Project success factors

Projects are temporary endeavors with their termination planned from the beginning. Thus, the project objectives and success criteria should be agreed upon with the stakeholders before starting a project [14, 15]. The long-term orientation needed to consider passive stakeholders contradicts the temporality of projects unless the long-term perspective is considered in the project objectives, business case, and investments [15]. Project success refers to the project delivering its expected output and achieving its intended objective. In contrast, project efficiency refers to the project management success regarding time, costs, and quality—the iron triangle [9, 14]. Success criteria and success factors are the dimensions for the stakeholder perceptions of project success [9, 14]. The criteria measure success while factors identify the circumstances, conditions, and events for reaching the objectives. The efficiency of the project can be measured when the outputs are produced. In contrast, project benefits and organizational performance impacts can be measured after the project outputs have been put into operations.

Several project management critical success factor models exist, the Pinto and Slevin [16] model being the most referenced [14]. It defines ten success factors under the control of the project team and four factors that influence project success but are not under the project team's control. Rather than identifying specific project success factors, [17] identified four groups of interrelated factors that could be analyzed across any type of project. While each model considers internal and external factors, their scope is bound by the project objectives. Customer consultation and acceptance are success factors; however, the public is not foreseen in that role. Using the framework from [16], [18] identified ethical knowledge as a specialized skill needed by the project personnel, questioned the role of moral decision-making in extreme situations, and identified ethical concerns as a risk to manage. However, the study does not explicitly address morality as a project objective.

The [9] project success model examines how stakeholders perceive success after completing the project. It was the first model to look at success outside the typical project life cycle and simultaneously consider multiple stakeholders [11]. The model defines the project results at multiple timescales: project outputs at the end of the project, outcomes months after the end, and impacts years after the end. It considers eight stakeholder groups: investors or owners, consumers, operators/users, project executive or project sponsor, project manager and project team, senior supplier, other suppliers, and public. Each stakeholder group and timescale provide success indicators, such as cost, features, performance, benefits, documentation, training, retention, well-being, learning, profit, new capabilities, future work, and new competence.

### B. AI projects

AI encompasses multiple disciplines or branches within computer science. Natural Language Processing (NLP) covers making the computer understand, process, and manipulate human language [19]. Pattern recognition is focused on classifying data into classes based on specific attributes [19]. Machine learning and deep learning are techniques used to define algorithms, and each uses data to learn [18-20]. Machine learning uses supervised and unsupervised methods to discover and model the patterns and relationships in data, allowing it to make predictions. Deep learning uses machine-learning approaches to automatically learn and extract features from complex unverified data without human involvement [18, 20]. Artificial neural networks, conceptually inspired by how the human brain works using biological neurons, are models trained on past data to make predictions [19]. The degree of human intervention in the decision-making process varies according to the type and purpose of the integration [18]. Technologies such as big data, predictive analytics, business intelligence, advanced analytics, and some digitization projects provide the foundation for these solutions. The technical processes for building algorithms require high-performance computing systems and architectures [18].

### C. Algorithmic decision-making

Algorithmic decision-making can be viewed as having three stages: development, usage, and consequence [8, 18]. The development stage produces an algorithmic system in three steps. The source data are collected from multiple sources; the data are made fit for purpose, including augmenting it with tags, identifiers, or metadata; and stored in data repositories. For the second step, subsets of source data are transformed into data for training the models (referred to as training data). The models and algorithms are developed through the extensive use of data and analytical methods. This activity is training the model. Here high-performance computing is needed to support the computational load and data volumes. The algorithms are validated. A user interface is developed for producing autonomous decisions or providing input for human decision-making. This step may include other technical aspects, such as system deployment; these topics are relevant but not the main focus of this study. In the usage stage, the algorithms are used by inputting parameters or data to invoke them; the algorithms output the decisions. The algorithm or systems may be standalone systems, integrated into other systems, robots, automobiles, etc., or exists in a digital technology platform such as a social media platform. In the consequences stage, the decision is finalized, and the consequences are realized on people, organizations, and groups.

### D. Morality and ethics in AI

Jones [21] defines a moral issue as one where a person's actions, when freely performed, has consequences (harms

or benefits) on others. The moral issue must involve a choice on the part of the actor or the decision-maker. He summarizes that many decisions have a moral component as they affect others. A moral agent is a person that makes the decision even when the decision-maker may not recognize a moral issue is at stake. An ethical decision is both legally and morally acceptable to the larger community; an unethical decision violates either the legal or the moral acceptability. Much of the research reviewed, treat the terms moral and ethical as equivalent and use them interchangeably depending on the context.

The thesis from [22] on morality is that the concepts of right and wrong should be discarded and replaced with a definition of morality in terms of “intrinsically unjust” versus “unjust given the circumstances.” He argues that the boundary between the two concepts is “according to what’s reasonable.” Anscombe [22] further theorizes that determining the expected consequences plays a part in determining what is just. These arguments place the responsibility for morality on the decision-maker. However, they do not answer who is accountable when the decisions are delegated from humans to systems.

Manders-Huits [7] argues that the notion of consequences and level of autonomy of action are preconditions or considerations for moral responsibility. First, the notion of consequences in information technology (IT) places the burden of responsibility for moral decisions on the shoulders of the designers of complex IT systems. However, the definition of the designers is unclear—technicians, finance providers, instructors—as well as how the designer’s responsibility relates to the responsibility the end users have for final decision-making. Martin [8] also places the responsibility for moral decision-making with the system developer and their companies. Second, the abundance of information that individuals have and understand enhances their possibility of action autonomy. The actions or decisions integrated into IT applications are limited based on “implying an adequate understanding of all relevant propositions or statements that correctly describe the nature of the action and the foreseeable consequences of the action” [7]. It is not likely that modelers can predict all potential uses of their models [23]. Consequently, [24] argues that machines are artificial agents that should not be held to a higher moral standard than humans and define five meta-moral qualities that machines should possess to be considered proper moral agents (robustness, consistency, universality, and simplicity).

A significant amount of research has focused on defining values, principles, frameworks, and guidelines for ethical AI development and deployment [13, 25]. However, [13] determined that principles alone have a limited impact on AI design and governance. Conducting an analysis of 21 AI ethic guidelines, [26] similarly found that AI guidelines are ineffective and do not change the behavior of

professionals from the technology community. One challenge is the difficulty in translating concepts, theories, and values into practice. Specifically, the translation process is likely to “encounter incommensurable moral norms and frameworks which present true moral dilemmas that principles cannot resolve” [13]. Furthermore, there are no proven methods to translate the principles into practice. Mittelstadt [13] warns that the solution to AI ethics should not be oversimplified to addressing only the AI technical design or expertise.

Jobin, et al. [25] conducted a content analysis of 84 AI ethical guidelines and identified five ethical principles that converged globally (transparency, justice, fairness, non-maleficence, and privacy). Building on the research from [25], [27] provides a detailed explanation of the normative implication of AI ethics guidelines for developers and organizational users. The paper provides a deep dive into the details. It specifies AI ethical principles and what users and developers ought to do to realize their moral responsibilities. However, the study explicitly excludes other stakeholders. Furthermore, in providing AI ethics research, [28] identified that AI ethics interests change over time.

### III. RESEARCH METHODOLOGY

This section describes the research methodology, including the theoretical model.

#### A. Theoretical Framework

To answer the research question, this research seeks the deliverables, acts or situations necessary to avoid harm or ensure benefits of an algorithm developed in projects. Thus, the project success model from [9] is relevant for identifying the success factors. It attempts to forecast project success beyond just the initial project outputs. It recognizes multiple stakeholders interested in the project output, outcomes, impacts, and that stakeholder interest change over time.

The model from [9] was chosen for four key reasons. First, the model focuses on projects and projects are bound by time, team, tasks, and activity. These boundaries limit environmental considerations. This is relevant as personal experience, organizational norms, industry norms, and cultural norms affect stakeholders’ perceived alternatives, consequences, and importance. Second, decisions made during the project will have an impact many months or years in the future. However, the project participants may not be aware of the magnitude of the consequences of their decisions in terms of harms or benefits on their victims or beneficiaries at the time of the decision. Thus, it is important to consider the multiple time dimension available in the model. Third, stakeholders influence the project’s planning and outputs and are impacted by the project results. Thus, the multiple stakeholder perspectives are useful for considering the influence of the decision-making

and the impact of the decisions on the stakeholders. Finally, the model outlines the multiple types of success indicators that should be considered in the investigation.

The algorithmic development, usages, and consequence stages and AI components were aligned with the timescales with the model from [9]. Algorithm development aligns with the project output, algorithm usage with the outcome, and decision-making consequences with the impact. Table I identifies the alignment of AI components to the time scales.

### B. Systematic Review Procedure

A systematic review of the literature was used to explore the research question. “A systematic review is a review of a clearly formulated question that uses systematic and explicit methods to identify, select, and critically appraise relevant research, and to collect and analyze data from the studies that are included in the review” [29]. The purpose of the systematic review was to synthesize existing knowledge in a structured and rigorous manner. The procedure included an 1) identification of bibliographic databases from which to collect the literature, 2) definition of the search process including the keyword and the search string, 3) definition of inclusions and exclusion criteria, 4) removing duplicates and screening the articles, 5) extracting data based on a full-text review of the articles, and 6) synthesizing the data using a coherent coding method. Details are described in the following sections, and Fig. 1 includes a flow of information through the systematic review. The process was conducted by a single researcher.

#### I. Bibliographic databases

The first literature search was in October 2020 for peer-reviewed articles in the ProQuest, Emerald, ScienceDirect, and IEEE Xplore bibliographic databases. The focal keywords were “algorithm” and “stakeholder.” This search revealed key themes in how success was viewed in algorithmic projects. Keywords such as ethics, fairness, accountability, transparency, and explainability were frequently referenced in the articles. The analysis identified the “ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT)” as an important source for cross-disciplinary research. Thus, bibliographic searches were undertaken in March and July 2021, adding the ACM Digital Library to the previous bibliographic databases.

#### II. Search string

The ultimate search was performed, placing emphasis on “accountability” instead of “stakeholder.” Stakeholder in combination with algorithm was not a frequent keyword, and accountability focuses on the relationship between project actors and those to whom the actors should be accountable [30]. Other frequent keywords were also included in the search string to make the results meaningful. Since not all databases allowed wild cards, variations of the search string were used, and adjustments were made in

TABLE I.  
PROJECT TIMESCALES AND AI COMPONENTS

Time scales	AI component
Output	<ul style="list-style-type: none"> <li>• Source Data, Data Collection and Storage</li> <li>• Training Data</li> <li>• Model and Algorithm Development</li> <li>• Model and Algorithm Validation</li> <li>• User Interface</li> <li>• System Architecture &amp; Configuration</li> </ul>
Outcome	<ul style="list-style-type: none"> <li>• Input Interface</li> <li>• Model and Algorithm Usage</li> </ul>
Impact	<ul style="list-style-type: none"> <li>• Decisions</li> </ul>

the syntax for each search engine. The wildcard version of the search string is as follows.

All=accountabl\* AND

Title = ("machine learning" OR "artificial intelligence" OR AI OR "big data" OR algorithm\*) AND

Title = (fair\* OR ethic\* OR moral\* OR success OR transparency OR explainabl\*)

#### III. Inclusions and exclusion criteria

Articles were retained in the search result for peer-reviewed journal articles or conference papers and English language; book reviews were excluded. There were no filters for the dates. Duplicate entries and entries with no document were removed. Next, literature was excluded or retained in an iterative process based on first a review of the title, second a review of the abstract, and finally a review of the full article text.

First, the title of the articles was reviewed and articles were retained that were about the process or considerations for the development, use, or outcomes of algorithms. Articles were excluded that were about the structure or content of an algorithm, a specific use case, or wrongly identified articles, e.g., magazine articles, panel descriptions. Next, the abstracts were reviewed to determine if the article could yield information on the success factors. Finally, the full text of the included articles was reviewed and coded to answer the research question. New articles identified during the analysis process were manually added. In total, the full-text of 85 articles were included for analysis. The majority of the included articles (79%) were published since 2019 and many (36%) were conference papers. Table II shows the article distribution by database and Fig. 1 shows the preferred reporting items for systematic reviews and meta-analyses (PRISMA) process flow.

#### IV. Data analysis

Each of the 85 articles was reviewed in detail for coding the success factors. The coding was conducted in Nvivo 12 (Windows) software. We extracted terms and explanations to determine what was known about how different stakeholders viewed success; we used the literature to clarify the definitions, provide examples, determine the main elements of success, and develop context. We compiled a resulting list of success factors that had to be deliverable, acts or situations that contributed to a positive outcome with

algorithm decision-making projects. The success factors were qualitatively grouped based on their common characteristics or responsibility patterns. The results are summarized in the Research Findings section.

### C. Validity and Reliability

This approach of defining elements is an acceptable method for placing boundaries around the meaning of a term [29]. First, internal consistency was provided by using a theoretical model to conduct the literature search and produce the guiding questions. Second, external validity was ensured by using literature as a primary source and a validation source. The success factors were mapped at a detailed level to their original sources in the literature. The results were cross-validated with prior AI ethic literature reviews from [27] to ensure completeness. The checklist and phase flow from the PRISMA Statement were used to guide the study and report the results [29].

## IV. RESEARCH FINDINGS

The literature review identified 71 success factors that were qualitatively consolidated in an iterative process into three broad categories and 14 groups. The results describe the practical requirements for success with AI development and usage based on the moral issues and ethical principles found in the literature. From an AI development perspective, the factors align with each of the AI components and translate principles into design and development requirements. For operations, the factors are the procedures for the usages of the algorithms and for addressing the concerns and expectations of the stakeholders. From a project point of view, the factors are the management concerns of end users, project sponsors, and project managers. Thus, the principles of trustworthiness, transparency, explainability, accountability, sustainability, etc., are distributed throughout the individual success factors.

First, features, capabilities or content of the deliverables were categorized as belonging to product qualities. Expectations, processes, or procedures related to the content and usage of the deliverables are placed in a procedure category. The third category of success factors relates to the management process, benefits, or protections expected.

TABLE II.  
ARTICLE DISTRIBUTION BY DATABASE

Database	Search Results	Duplicate	Screened by		Eligible
			Title	Abstract	
ACM	172	10	162	139	23
Emerald	8	4	4	2	1
IEEE Xplore	117	8	109	20	10
ProQuest Science Direct	118	14	104	96	31
Manual	20	2	72	16	6
	509	38	471	282	85

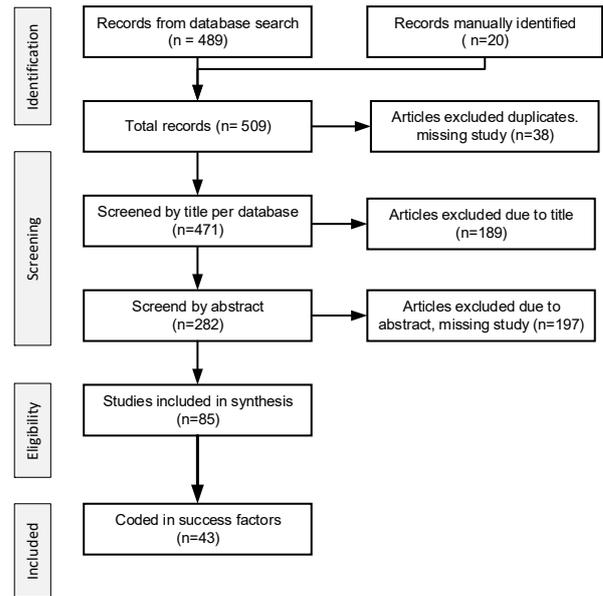


Fig 1. PRISMA process flow

These categories align with the conduct groups in the accountability model referenced in [30].

The characteristics of the factors or their impact or influence by project actors or stakeholders influenced the categorization. Table III identifies the success factors based on their categories and groups. This section describes each success factor by category and group; the factors are italicized in the text.

### A. Product Qualities

#### 1. Source Data Qualities

*Data accessibility* refers to access and use of data in the algorithm creation process. Several regulations and laws constrain how data may be accessed, processed, and used in analytical processes. Thus, a legal agreement to use the data and confidentiality of personal data should be preserved [1, 3, 31-35]. *Data transparency* refers to revealing the source of the data collected, including the context or purpose of the data collection, the application, the sensors (or users that collected the data), and the location in which the data are stored [23, 33, 36-40]. The reviewability framework [41] recommends maintaining *data collection records* of data and their lifecycle. The recommended content includes providing details on purpose, creators, funders, composition, content, collection process, usage, distribution, limitations, maintenance, and data protection and privacy concerns [2, 33, 34, 36, 38, 41]. Datasheet by [36] provides detailed guidance on document content.

#### II. Training Data Qualities

In interacting with and processing data, individuals are entitled to physical and psychological safety, i.e., *interaction safety* [3, 23, 27, 33, 35, 42, 43].

TABLE III.  
SUCCESS GROUPS AND SUCCESS FACTORS

Category	Success Groups	Success Factors	References
Product Qualities	Source Data Qualities	Data accessibility, Data transparency, Data collection records	[1-3, 23, 31-41]
	Training Data Qualities	Data quality and relevance, Interaction safety, Equitable representation, Model training records	[3, 23, 27, 33, 35, 42-46]
	Models & Algorithms Qualities	Algorithm transparency, Consistency, Accuracy, Interpretability, Auditability, Model validation, Algorithm renewal, Model validation records	[1-3, 23, 27, 31, 33, 37, 38, 45-52]
	User Interface Qualities	Human intervention, Equitable accessibility, Front-end transparency	[1, 23, 27, 31, 32, 34, 37, 39, 53, 54]
	System Configuration	System and architecture quality, Security safeguards, Technical logging, Technical deployment records	[2, 3, 31, 41, 49, 55]
	Data Privacy & Confidentiality	Informed consent, Personal data controls, Confidentiality, Privacy safeguards, Data anonymization, Data encryption, Data retention policy	[1, 3, 23, 27, 31-33, 37]
Procedures	Decision Quality	Awareness, Access and redress, Decision accountability, Equitable treatment, Privacy and confidentiality, Civil rights and liberty protections	[1-3, 23, 27, 32, 34, 37, 40, 42, 44, 49, 56]
	System Transparency & Understandability	User-centric communication, Interpretable models, Choices, Specialized skills and knowledge, Interaction safety, Problem reporting, Usage records	[2, 23, 27, 31, 33-35, 37, 40, 41, 49, 57-59]
	Usage Controls	Compliant process, Quality controls, Monitoring, Consequence records, Process deployment records	[23, 34, 39, 41, 51, 59]
	Investigation	Algorithm auditing, Audit finding records, Audit response records, Algorithm impact assessments, Certification	[1-3, 23, 30, 41, 47, 51, 60]
Management	Governance	Scope definition document, Responsibility assignment matrix, Diverse working environment, Ethics policies, Recordkeeping, Risk assessment records, Disclosure records, Procurement records	[10, 27, 34, 41, 61-63]
	Financial Benefits	Intellectual property rights, Profits, License or service fees, Investment funds	[23, 40, 42]
	Financial Protections	Intellectual property protection, Environmental impacts, Energy costs, Cost efficiency, Project efficiency	[9, 14, 27, 42, 46, 47]
	Legal Protections	Limiting liability, Legal safeguards, Regulatory and legal compliance	[2, 3, 23, 27, 31, 33, 40, 41, 51, 60]

*Equitable representation* applies to data and people. For data, it means having enough data to represent the whole population for whom the algorithm is being developed while also considering the needs of minority groups such as handicapped people, minors (under 13 years old), and ethnic minorities. For people, it means, for example, including representatives from minority groups or their advocates in the project governance structures or teams that design and develop algorithms [23, 44-46]. *Model training records* should document the training work flow, model approaches, predictors, variables, and other factors; datasheets by [36] and model cards by [48] provide a framework for the documentation.

### III. Model & Algorithm Qualities

*Algorithm transparency* refers to using straightforward language to provide clear, easily accessible descriptive information (including trade secrets) about the algorithms and data and explanations for why specific recommendations or decisions are relevant. The need for end users to understand and explain the decisions produced by the algorithms determines the algorithm, data, and user interface transparency requirements [1, 23, 31, 33, 37, 47, 48]. Model qualities include consistency, accuracy, interpretability, and suitability; there are no legal standards for acceptable error rates or ethical designs. *Consistency* means receiving the same results given the same inputs; nondeterministic effects can occur based on architectures with opaque encodings or imperfect computing environments [3]. *Accuracy* is how effective the model provides the desired output with the fewest mistakes (e.g., false positives,

error rates) [3, 23, 37, 45, 46]. *Interpretability* refers to the degree to which the model is designed to provide reliable and easy-to-understand explanations of its prediction [27, 37, 49]. *Auditability* refers to how the algorithm is transparent to or obfuscated from an external view to allow other parties to monitor or critique it [2, 38].

*Model validation* is the execution of mechanisms to measure or validate the models for adherence to defined principles and standards, effectiveness, performance in typical and adverse situations, and sensitivity. The validation should include bias testing, i.e., an explicit attempt to identify unfair bias, avoid individual and societal bias, and reverse any biases detected. Models can be biased based on a lack of representations in the training data or how the model makes decisions, e.g., the selected input variables. The model outcomes should be traceable back to input characteristics [2, 23, 50-52]. Model values or choices become obsolete. They need to be reviewed or refreshed so an *algorithm renewal* process should be established [23, 30]. The reviewability framework suggests maintaining *model validation records* that contain details on and how the model was validated, including dates, version, intended use, factors, metrics, evaluation data, training data, quantitative analyses, ethical considerations, caveats and recommendations, or any other restrictions [41, 48]. Model cards by [48] provide detailed guidance on the content.

### IV. User Interface Qualities

Expertise is embodied in a model in a generalized form that may not be applicable in individual situations. Thus,

*human intervention* is the ability to override default decisions [1, 34, 37]. *Equitable accessibility* ensures usability for all potential users, including people with disabilities [23, 27, 53]. *Front-end transparency* designs should meet transparency requirements and not unduly influence, manipulate, confuse, or trick users [31, 32, 39, 54]. Furthermore, dynamic settings or parameters should consider context to avoid individual and societal biases such as those created by socio-demographic variables [34]. App-Synopsis by [54] provides detailed guidance on the content.

#### V. System Configuration

The *system and architecture quality* may impact the algorithm's outcomes, introduce bias, or result in indeterminate behavior. Default choices (e.g., where thresholds are set and the defaults to be specified) may introduce bias in the decision-making. Specifically, the selected defaults may be based on the personal values of the developer. Decisions on methods and the parallelism of processes may cause system behavior that does not always produce the same results when given the same inputs. Obfuscated encodings may make it difficult to process the results or audit the system. The degree of automation may limit the user's choices [3, 49]. *Security safeguards* are implementing technology, processes, and people to resist accidental, unlawful, or malicious actions that compromise the availability, authenticity, integrity, and confidentiality of data [2, 31, 55].

The reviewability framework suggests the systems should provide a *technical logging* process including mechanisms to capture the details of inputs, outputs, and data processing/computation. The framework also recommends records relevant to the *technical deployment records* and operations, including installation procedures, hardware, software, network, storage provisions or architectural plans, system integration, security plans, logging mechanisms, technical audit procedures, technical support processes, maintenance procedures [41].

#### B. Procedures

##### 1. Data Protection, Privacy and Confidentiality

*Informed consent* is the data subject's right to be informed on the collection, use, and repurposing of their personal data [3, 23, 27, 31, 37]. The legal and regulatory rules covering consent vary by region and usage purposes. *Personal data control* means giving people control of their personal data [1, 32, 37]. *Confidentiality* concerns protecting and keeping confidential data and proprietary information. *Privacy safeguards* include processes, strategies, guidelines, and measures to protect and safeguard data privacy, along with remedies for privacy breaches. For example, a privacy measure could be data encryption or data anonymization [1, 3, 23, 32, 33, 37]. *Data anonymization* involves applying rules and processes that randomize data so an individual is not personally identifiable and cannot

be re-identified through combining data sources. In general, data protection principles do not apply to anonymous information [23, 32, 33]. *Data encryption* is an engineering approach to secure data with electronic keys. *Data retention policy* specifies the time and obligations for keeping data [31].

##### II. Decision Quality

*Awareness* is educating the public about the existence and the degree of automation, the underlying mechanisms, and the consequences [2, 37]. *Access and redress* are a way to investigate and correct erroneous decisions. It includes the ability to contest automated decisions, including expressing a point-of-view or requesting human intervention in the decision [1, 2, 27, 37, 40, 56]. *Decision accountability* is knowing who is accountable for the actions when decisions are taken by the automated systems in which the algorithms are embedded [2, 27, 56]. *Equitable treatment* means eliminating discrimination and differential treatment, whereby similarly situated people are given similar treatment. In this context, discrimination does not only equate to prejudice based on race. It is based on forming groups using 'statistical discrimination'; it further refers to anti-discrimination and human rights protections [1, 2, 27, 34, 42, 44]. *Privacy and confidentiality* are the activities for protecting and keeping confidential information of an identified or identifiable natural person [3, 23, 27, 34, 40, 42, 49]. In this context, *civil rights and liberties protection* are securing and providing the fundamental rights and freedoms of natural persons, including the right to data protection and privacy and to have opinions and decisions made independently of an automated system [27, 32].

##### III. System Transparency & Understandability

*User-centric communication* considers the explainability of the algorithm to the intended audience. It transmits essential, understandable information rather than legalistic terms and conditions. Explanations are communicated in layman's terms, even for complex algorithms [2, 27, 31, 34, 37, 40, 57]. *Interpretable models* refer to having a model design that is reliable, understandable, and possible for expert users to explain the predictions [27, 37]. *Choices* allow users to decide what to do with model results or, in other words, provides a degree of human control [23, 27, 37, 49, 58].

Expertise is embodied in a generalized form that may not be applicable in individual situations, so *specialized skills and knowledge* may be required to choose between alternatives. Consequently, professional expertise, staff training and supervision, and on-the-job coaching may be necessary to ensure appropriate use and decision quality [49, 59].

*Interaction safety* refers to ensuring physical and psychological safety for the people interacting with the AI systems [35]. *Problem reporting* is a mechanism that allows

users to discuss and report concerns such as bugs or algorithmic biases [47]. The reviewability framework recommends retaining *usage records* of model inputs and outputs of parameters, operational records at the technical (systems log) level, usage instructions [33, 41].

#### IV. Usage Controls

The *complaint process* means having mechanisms in place for identifying, investigating, and resolving improper activity or receiving and mediating complaints [39]. *Quality controls* detect improper usage or under-performance. Improper usage occurs when the system is used in a situation for which it was not originally intended [23, 34]. *Monitoring* is a continual process of surveying the system's performance, environment, and staff for problem identification and learning [59]. System monitoring is to verify how the system behaves in unexpected situations and environments. The staff monitoring identifies absent or inadequate content areas, identifies systematic errors, anticipates and prevents bias, and identifies learning opportunities.

The reviewability framework recommends retaining consequence and process deployment records. *Consequence records* document the quality assurance processes for a decision and log any actions taken to affect the decision, including failures or near misses [41, 51]. Logging and recording decision-making information are appropriate means of providing traceability. *Process deployment records* document relevant operational and business processes, including workflows, operating procedures, manuals, staff training and procedures, decision matrices, operational support, maintenance processes and records [41].

#### V. Investigations

*Algorithm auditing* is seen as a method for understanding how algorithms work. Testing algorithms based on issues that should not arise and making inferences from the algorithms' data is a technique for auditing complex algorithms [1-3, 41, 47, 51]. Audit records include audit finding records and audit response records. *Audit finding records* document the audit, the basis or other reasons it was undertaken, how it is conducted, who conducted it, any findings [41]. *Audit response records* document remediations and subsequent actions or remedial responses based on audit findings [2, 41].

*Algorithmic impact assessments* investigate aspects of the system to render visible impacts of the systems and propose steps to address any deficiencies or harms [30, 60]. *Certification* identifies that people or institutions comply with regulations and safeguards and publicize institutions with breaches; it offers independent oversight by an external organization [23, 51].

### C. Management

#### I. Governance

The *scope definition document*, or problem statement, defines the aims and rationale for the algorithmic system [10, 41]. The requirement for the system, the moral issues,

and all aspects of the project are impacted by the context (country, industry sector, functional topic, and use case) of the algorithm. Trust is context-dependent since things can work in one context but not another; thus, the scope should act as a contract that makes explicit the algorithm's goal and the behavior that can be anticipated [61]. Furthermore, a clearly defined scope protects against spurious claims and misapplication or misuse of the system. Next, ethical principles argue AI systems should be developed to do good or benefit someone or the society as a whole (beneficence); they should avoid doing harm to others (non-maleficence) [27, 34]. Finally, rules should be established on managing conflict of interest situations within the team or when the values of the system conflict with the interests or values of the users [62, 63].

A *responsibility assignment matrix* defines roles and responsibilities within a project organization. It distinguishes between persons or organizations with responsibility and accountability [64]; accountability ensures a task is satisfactorily done, and responsibility accepts an obligation to perform a task satisfactorily, with transparency in reporting on outcomes, corrective actions, or interactive controls [64, 65]. Both responsibility and accountability assume a degree of subject matter understanding and knowledge [27]. The project organization should promote a *diverse working environment*, including involving various stakeholders and people from differing backgrounds and disciplines and promoting the exchange and cooperation across regions and between organizations [27, 43].

*Ethics policies* should include guidelines and rules for implementing, verifying, and remedying ethical principles; the guides should be shareable externally with the public or public authorities [2, 27, 33]. The practical aspects of ethical principles for fairness, trustworthiness, transparency, explainability, accountability, and sustainability are distributed throughout the individual success factors discussed in this study.

Systematic *recordkeeping* is the mechanism for retaining logs and other documents of contextual information about the process, decisions, and decision-making from the project inception through the system operations [10, 27, 33, 41, 49]; the various types of records are recorded as individual success factors. The *risk assessment records* identify the potential implications and risks of the system such as legality and compliance, discrimination and equality, impacts on basic rights, ethical issues, sustainability concerns [10, 41]; *disclosure records* are logs that are themselves about disclosures or the processes for disclosure, what was actually released, how information was compiled, how it was delivered, in what format, to whom, and when [31, 33, 41]; and *procurement records* are contractual arrangements, tender documents, design specifications, quality assurance measures, and other documents that detail the suppliers and relevant due-diligence [41].

## II. Financial Benefits

*Intellectual property rights* consist of the ownership of the design of the models, including the indicators. Innovation levels have to be balanced with risks of liabilities and litigation for novel concepts [23]. *Profits* include increased revenues from the sale or licensing models that produce revenue through *license or service fees* [23, 42] or reductions in costs from making faster, less expensive, or better decisions [40]. Furthermore, proven successful models, concepts, algorithms, or business models can attract *investment funds* [23].

## III. Financial Protections

*Intellectual property protection* is achieved by partly or entirely hiding the algorithm's design choices. Data and algorithm transparency and auditing requirements should be considered in deciding what to reveal [47]. Model development has environmental impacts and energy costs. The *environmental impacts* occur as the big training models may be energy-intense using as much computing energy as a trans-American flight in carbon emissions [27, 46]. The *energy costs* from computing power and electricity consumption (for on-premise or cloud-based services) are relevant for training models [27, 46]; for an incremental increase in accuracy, the cost of training a single model may be extreme (e.g., 0.1 increase in accuracy for 150,000 USD) [46]. *Cost efficiency* occurs acquiring and using information is less than the costs involved if the data were absent [42]. *Project efficiency* evaluates the project management's success in meeting stakeholder requirements for quality, schedule, and budget [9, 14].

## IV. Legal Protections

The *legal safeguards* include protection against legal claims or regulatory issues that arise from algorithmic decisions [2, 31]. *Limiting liability* or risk of litigation for users and balancing risks from adaptations and customizations with fear of penalties or liability in situations of malfunction, error or harms [23, 27]. *Regulatory and legal compliance* involves meeting the legal and regulatory obligations for collecting, storing, using, processing, profiling, and releasing data or complying with other laws, regulations, or ordinances [3, 33, 40, 41, 51, 60].

## V. DISCUSSION

This study framed the question of project success from the perspective of moral decision-making with algorithms. People impacted by algorithm decisions want fairness, meaning moral or "just" treatment from algorithmic decision-making. However, fairness or the perception of fairness has several subjective components that are out of the scope of any development project, including pre-established attitudes and emotional reactions to algorithmic outcomes [1, 44]. Moreover, [66] empirically found that end users understand, perceive, and process algorithm fairness, accountability, and transparency differently. Furthermore,

the interaction between trust and algorithmic features influence user satisfaction.

Nevertheless, the research revealed that the project team's actions influence who is the judge of what is reasonable when the decision is made [7]. Thus, the project team has some responsibility for the moral decisions produced by the algorithmic systems. The limits and bias in decisions produced, the end user's ability to manipulate the system or override the decisions, and the information the end users have to understand and enhance their decision autonomy mediate the project team's accountability. The project organization can take on some responsibility by considering moral decision-making in the project scope. The success factors for the product qualities, procedures, and management are discussed in the following sections.

### A. Product Qualities

The product quality success factors must be considered from many external stakeholders' perspectives, including individuals, society, end users, user organizations, technology platforms, etc. Thus, each development aspect needs to consider the technical product qualities, usability features, information requirements, and legal and regulatory requirements. In this regard, several conflicting success factors have to be balanced. For example:

- The end users may want a high degree of flexibility for human intervention, including making alternative choices. Similarly, the person impacted by the decision outcome will want to have erroneous (or biased) decisions reviewed and corrected. Conversely, the user's organization would want to limit legal liabilities, which speaks for fewer choices. The more open the system, the harder it is to differentiate between a system error and user error and assign accountability.
- The need for the end users to understand and explain the decisions produced by the algorithms suggests a high degree of transparency for the algorithm, data, and front-end user interface. Conversely, the need to preserve intellectual property rights is a factor for a lesser degree of transparency.
- Unbiased models can produce high error rates (or be inaccurate), and biased models can be accurate. Thus, there is a tradeoff between utility and fairness due to bias or inaccuracies.
- There is a tradeoff between the degree of automation and human autonomy. Too much automation can give the perception (or reality) that people are under constant surveillance or that the system knows too much and is what [40] calls creepy. Meanwhile, the system can offer flexibility, accuracy, or benefits not available through human autonomy.
- Developing large-scale language models produces carbon emissions and has a financial cost. However, the assumption (which is challenged) is that large models

increase accuracy. Thus, there is a tradeoff between accuracy, environmental impacts, and financial costs.

The success factors in the data protection, privacy and confidentiality group relate to the product qualities and usage procedures. Specific product capabilities may be needed to realize certain processes. For example, personal data controls require some degree of system traceability and extracts for personal data. Also, it must be clear when data stops being personal within the system and becomes generic or anonymized. Another overlap between product qualities and procedures relates to the methods and practices used to implement privacy controls or data anonymization.

#### B. Procedures

The procedures for using and investigating algorithmic systems depend on the many product qualities and procedures enabled by the system or implemented by the end users and their organization. The user's organization and the platform providers must follow regulations and laws relevant to the industry, data processing, and data profiling. Furthermore, as of April 2021 in the European Union, the artificial intelligence regulation act requirements should be considered [67]. Thus, success factors are robust operational rules, policies, contracts; quality controls; and privacy and security safeguards.

#### C. Management

There are several success factors from a business and governance perspective for delivering the product, intellectual property rights and protections, limiting liability, ensuring legal safeguards and regulatory compliance. Similar to the product qualities, there are multiple conflicting success factors. For example:

- The tradeoff between accuracy, environmental costs, and financial costs is already discussed.
- The need for financial profits from algorithm systems and the need to benefit society (beneficence) may result in conflicting objectives.
- The project efficiency concerning quality, time, and budget and the regulatory and legal compliance.
- The need for algorithm, data, and front-end user interface transparency and producing intellectual property rights and protections.
- The need for legal safeguards, comparing the need for system flexibility to allow for choices at the point of decision versus restricting human intervention.

## VI. CONCLUSIONS

The importance of algorithms in society and individuals' lives is becoming increasingly apparent. Therefore, the success factors for AI projects are important and are dramatically more expansive than those for a typical information systems project. This research identified 71 AI project success factors in 14 groups related to moral decision-

making with AI projects. The research summarizes the concerns for fair, moral algorithm development and usage in decision-making. It reveals the project manager and the project team need to consider many factors when defining the project scope and executing it. An AI development project has a narrower scope and fewer short-term implications than an algorithmic development project considering moral decision-making. This paper argues that the people that develop and operate AI systems are moral agents. Those people should build AI systems and procedures to avoid harm and ensure benefits. Hence, as artificial agents, the systems should abide by the moral laws of society.

Projects are constrained by time and budget, limiting the availability of people and other resources. Nevertheless, the importance of the algorithms that result from AI projects can be significant. Thus, it is necessary and relevant that a broad view of success factors be considered in planning and executing these projects. The findings from this study provide some guidelines on the success factors that may only be used indirectly or overtime to judge the project's success.

#### A. Practical Implications

Projects, and especially AI projects, are context-sensitive. The factors presented are generic; it would be important to adjust and validate in specific contexts. For example, developing an algorithm for a healthcare situation would have different considerations than an algorithm for a marketing situation.

The success factors provide insights into the activities and deliverables that should be considered part of the planning to ensure fair, ethical decisions. First, the project manager and sponsor should alter project scopes to consider moral decision-making with algorithms. This will dramatically affect the team compositions and the deliverables produced as part of the project. The benefits to society and the environment could be highlighted and potentially measured.

Next, project managers and sponsors may be limited in influencing future usage and operational processes. Nevertheless, they should try to exert this influence on the ethical practices of system users and user organizations. Furthermore, they should consider the success factors described herein to recognize moral issues that require decisions during the development process to mitigate project risks. Finally, as an agent of the sponsoring organization with a reputation to manage and business objectives to reach, the project manager should consider these success factors to ensure adherence to ethical, privacy, and security norms and deliver business benefits.

#### B. Theoretical Implications

The research expands the existing project management literature on project success factors specific to the AI domain. This contribution is consistent with the direction

identified by [14], "...we argue that one should turn to context-specific and even symbolic and rhetoric project success and CSFs [critical success factors]."

### C. Limitations and Future Research

This research was based on the latest available literature, but at a single point in time. AI is a fast-moving topic judging from the number of recent articles. Thus, other methods such as a Delphi study with field experts could extend and update the study and validate the findings. Since the analysis was conducted by a single researcher, the results may be biased by the researcher's perspective.

As an opportunity for additional research, the success factors could be used to investigate project accountability or stakeholder management. It could be expanded to identify measurable success criteria for some success factors. AI literature regarding ways to measure bias, inequality, and accuracy should be left to the specialists; however, it would be interesting to understand how to evaluate the tradeoffs needed during the projects and still meet all stakeholder requirements retaining an honest approach.

### REFERENCES

- [1] N. Helberger, T. Araujo, and C. H. de Vreese, "Who is the fairest of them all? Public attitudes and expectations regarding automated decision-making," *Computer Law & Security Review*, vol. 39, pp. 1–16, Nov 2020.
- [2] S. Garfinkel, J. Matthews, S. S. Shapiro, and J. M. Smith, "Toward algorithmic transparency and accountability," *Communications of the ACM*, vol. 60, p. 5, 2017.
- [3] J. A. Sherer, "When Is a Chair Not a Chair?: Big Data Algorithms, Disparate Impact, and Considerations of Modular Programming," *Computer and Internet Lawyer*, vol. 34, pp. 6–10, Aug 2017.
- [4] S. Baruffaldi, B. v. Beuzekom, H. Dernis, D. Harhoff, N. Rao, D. Rosenfeld, et al., "Identifying and measuring developments in artificial intelligence," 2020.
- [5] A. J. Shenhar, D. Dvir, O. Levy, and A. C. Maltz, "Project success: a multidimensional strategic concept," *Long range planning*, vol. 34, pp. 699–725, 2001.
- [6] S. J. Bennett, "Investigating the Role of Moral Decision-Making in Emerging Artificial Intelligence Technologies," presented at the Conference Companion Publication of the 2019 on Computer Supported Cooperative Work and Social Computing, Austin, TX, USA, 2019.
- [7] N. Manders-Huits, "Moral responsibility and IT for human enhancement," presented at the Proceedings of the 2006 ACM symposium on Applied computing, Dijon, France, 2006.
- [8] K. Martin, "Ethical Implications and Accountability of Algorithms," *Journal of Business Ethics*, vol. 160, pp. 835–850, Dec 2019.
- [9] R. J. Turner and R. Zolin, "Forecasting Success on Large Projects: Developing Reliable Scales to Predict Multiple Perspectives by Multiple Stakeholders Over Multiple Time Frames," *Project Management Journal*, vol. 43, pp. 87–99, 2012.
- [10] O. Zwikael and J. R. Meredith, "Who's who in the project zoo? The ten core project roles," *International Journal of Operations & Production Management*, vol. 38, pp. 474–492, 2018.
- [11] K. Davis, "An empirical investigation into different stakeholder groups perception of project success," *International Journal of Project Management*, vol. 35, pp. 604–617, May 2017.
- [12] R. K. Mitchell, B. R. Agle, and D. J. Wood, "Toward a Theory of Stakeholder Identification and Salience: Defining the Principle of who and What Really Counts," *Academy of Management Review*, vol. 22, pp. 853–886, Oct 1997.
- [13] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence*, vol. 1, pp. 501–507, 2019.
- [14] L. A. Ika, "Project success as a topic in project management journals," *Project Management Journal*, vol. 40, pp. 6–19, 2009.
- [15] C. Weninger, "Project Initiation and Sustainability Principles: What Global Project Management Standards Can Learn from Development Projects when Analyzing Investment," presented at the Paper presented at PMI® Research and Education Conference, Limerick, Munster, Ireland, 2012.
- [16] J. K. Pinto and D. P. Slevin, "Critical Success Factors Across the Project Life Cycle," *Project Management Journal*, vol. 19, p. 67, 1988.
- [17] W. Belassi and O. I. Tukel, "A new framework for determining critical success/failure factors in projects," *International Journal of Project Management*, vol. 14, pp. 141–151, 1996.
- [18] G. J. Miller, "A conceptual framework for interdisciplinary decision support project success," in *2019 IEEE Technology & Engineering Management Conference (TEMSCON)*, 2019, pp. 1–8.
- [19] J. Aggarwal and S. Kumar, "A Survey on Artificial Intelligence," *International Journal of Research in Engineering, Science and Management* vol. 1, Dec 2018.
- [20] R. Iqbal, F. Doctor, B. More, S. Mahmud, and U. Yousuf, "Big Data analytics and Computational Intelligence for Cyber-Physical Systems: Recent trends and state of the art applications," *Future Generation Computer Systems*, pp. 766–778, Nov 2017.
- [21] T. M. Jones, "Ethical decision making by individuals in organizations: An issue-contingent model," *Academy of management review*, vol. 16, pp. 366–395, 1991.
- [22] G. E. M. Anscombe, "Modern moral philosophy," *Philosophy*, vol. 33, pp. 1–19, 1958.
- [23] I. G. Cohen, R. Amarasingham, A. Shah, B. Xie, and B. Lo, "The Legal And Ethical Concerns That Arise From Using Complex Predictive Analytics In Health Care," *Health Affairs*, vol. 33, pp. 1139–1147, Jul 2014.
- [24] N. P. Shaw, A. Stöckel, R. W. Orr, T. F. Lidbetter, and R. Cohen, "Towards Provably Moral AI Agents in Bottom-up Learning Frameworks," *Aies '18*, pp. 271–277, 2018.
- [25] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, vol. 1, pp. 389–399, 2019.
- [26] T. Hagendorff, "The Ethics of AI Ethics: An Evaluation of Guidelines," *Minds and Machines*, vol. 30, Mar 2020.
- [27] M. Ryan and B. C. Stahl, "Artificial intelligence ethics guidelines for developers and users: clarifying their content and normative implications," *Journal of Information, Communication and Ethics in Society*, vol. 19, pp. 61–86, 2021.
- [28] Y. Zhang, M. Wu, G. Y. Tian, G. Zhang, and J. Lu, "Ethics and privacy of artificial intelligence: Understandings from bibliometrics," *Knowledge-Based Systems*, vol. 222, p. 106994, 2021/06/21/ 2021.
- [29] D. Moher, A. Liberati, J. Tetzlaff, and D. G. Altman, "Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement," *International Journal of Surgery*, vol. 8, pp. 336–341, Jan 2010.
- [30] M. Wieringa, "What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability," presented at the Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 2020.
- [31] A. Rossi and G. Lenzini, "Transparency by design in data-informed research: A collection of information design patterns," *Computer Law & Security Review*, vol. 37, pp. 1–22, Jul 2020.
- [32] M. Büchi, E. Fosch-Villaronga, C. Lutz, A. Tamò-Larrieux, S. Velidi, and S. Viljoen, "The chilling effects of algorithmic profiling: Mapping the issues," *Computer Law & Security Review*, vol. 36, pp. 1–15, Apr 2020.
- [33] E. Bertino, A. Kundu, and Z. Sura, "Data Transparency with Blockchain and AI Ethics," *Journal of Data and Information*

- Quality*, vol. 11, pp. 1–8, 2019.
- [34] M. Loi, C. Heitz, and M. Christen, "A Comparative Assessment and Synthesis of Twenty Ethics Codes on AI and Big Data," in *2020 7th Swiss Conference on Data Science (SDS), 26-26 June 2020*, 2020, pp. 41–46.
- [35] I. Munoko, H. L. Brown-Liburd, and M. Vasarhelyi, "The Ethical Implications of Using Artificial Intelligence in Auditing: JBE," *Journal of Business Ethics*, vol. 167, pp. 209–234, Nov 2020.
- [36] T. Gebru, J. Morgenstern, B. Vecchione, J. W. Vaughan, H. Wallach, H. Daumé III, *et al.*, "Datasheets for datasets: arXiv preprint arXiv:1803.09010," *arXiv preprint arXiv:1803.09010*, 2018.
- [37] R. Hamon, H. Junklewitz, G. Malgieri, P. De Hert, L. Beslay, and I. Sanchez, "Impossible Explanations? Beyond explainable AI in the GDPR from a COVID-19 use case scenario," presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 2021.
- [38] B. Hutchinson, A. Smart, A. Hanna, E. Denton, C. Greer, O. Kjartansson, *et al.*, "Towards Accountability for Machine Learning Datasets: Practices from Software Engineering and Infrastructure," presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 2021.
- [39] B. Wagner, K. Rozgonyi, M.-T. Sekwenz, J. Cobbe, and J. Singh, "Regulating transparency? Facebook, Twitter and the German Network Enforcement Act," presented at the Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 2020.
- [40] H. J. Watson and N. Conner, "Addressing the Growing Need for Algorithmic Transparency," *Communications of the Association for Information Systems*, vol. 45, p. 26, Mar 2019.
- [41] J. Cobbe, M. S. A. Lee, and J. Singh, "Reviewable Automated Decision-Making: A Framework for Accountable Algorithmic Systems," presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 2021.
- [42] O. H. Gandy, "Engaging rational discrimination: exploring reasons for placing regulatory constraints on decision support systems," *Ethics and Information Technology*, vol. 12, pp. 29–42, Mar 2010.
- [43] J. H. Lim and H. Y. Kwon, "A Study on the Modeling of Major Factors for the Principles of AI Ethics," presented at the Digital Government Research (DG.O '21), June 09–11, 2021, Omaha, NE, USA, 2021.
- [44] H. Adam, "The ghost in the legal machine: algorithmic governmentality, economy, and the practice of law," *Journal of Information, Communication and Ethics in Society*, vol. 16, pp. 16–31, 2018.
- [45] J. Alasadi, A. A. Hilli, and V. K. Singh, "Toward Fairness in Face Matching Algorithms," presented at the Proceedings of the 1st International Workshop on Fairness, Accountability, and Transparency in MultiMedia, Nice, France, 2019.
- [46] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?," presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Virtual Event, Canada, 2021.
- [47] M. Eslami, K. Vaccaro, M. K. Lee, A. E. B. On, E. Gilbert, and K. Karahalios, "User Attitudes towards Algorithmic Opacity and Transparency in Online Reviewing Platforms," presented at the Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, Scotland UK, 2019.
- [48] M. Mitchell, S. Wu, A. Zaldivar, P. Barnes, L. Vasserman, B. Hutchinson, *et al.*, "Model Cards for Model Reporting," presented at the Proceedings of the Conference on Fairness, Accountability, and Transparency, Atlanta, GA, USA, 2019.
- [49] M. Langer and R. N. Landers, "The future of artificial intelligence at work: A review on effects of decision automation and augmentation on workers targeted by algorithms and third-party observers," *Computers in Human Behavior*, vol. 123, p. 106878, 2021.
- [50] A. R. Givens and M. R. Morris, "Centering disability perspectives in algorithmic fairness, accountability and transparency," presented at the Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency, Barcelona, Spain, 2020.
- [51] U.-V. Albrecht, "Transparency of Health-Apps for Trust and Decision Making," *Journal of Medical Internet Research*, vol. 15, pp. 1–5, Dec 2013.
- [52] E. P. Vallejos, A. Koene, V. Portillo, L. Dowthwaite, and M. Cano, "Young People's Policy Recommendations on Algorithm Fairness," presented at the Proceedings of the 2017 ACM on Web Science Conference, Troy, New York, USA, 2017.
- [53] J. Matthews, "Patterns and Antipatterns, Principles, and Pitfalls: Accountability and Transparency in Artificial Intelligence," *AI Magazine*, vol. 41, pp. 82–89, Nov 2020.
- [54] U. Bhatt, A. Xiang, S. Sharma, A. Weller, A. Taly, Y. Jia, *et al.*, "Explainable Machine Learning in Deployment," presented at the Conference on Fairness, Accountability, and Transparency (Fat\* '20), January 27–30, 2020, Barcelona, Spain, 2020.
- [55] A. Mowbray, P. Chung, and G. Greenleaf, "Utilising AI in the legal assistance sector—Testing a role for legal information institutes," *Computer Law & Security Review*, vol. 38, pp. 1–9, Sep 2020.
- [56] A. Joerin, M. Rauws, R. Fulmer, and V. Black, "Ethical Artificial Intelligence for Digital Health Organizations," *Cureus*, vol. 12, May 2020.
- [57] B. Shneiderman, "Bridging the Gap Between Ethics and Practice: Guidelines for Reliable, Safe, and Trustworthy Human-Centered AI Systems," *ACM Trans. Interact. Intell. Syst.*, vol. 10, 2020.
- [58] J. Metcalf, E. Moss, E. A. Watkins, R. Singh, and M. C. Elish, "Algorithmic Impact Assessments and Accountability: The Co-construction of Impacts," presented at the Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, 2021.
- [59] A. Jacovi and Marasovi, "Formalizing Trust in Artificial Intelligence: Prerequisites, Causes and Goals of Human Trust in AI," presented at the ACM Conference on Fairness, Accountability, and Transparency (ACM FAccT '21), March 3–10, 2021, Virtual Event, Canada, 2021.
- [60] A. Aguirre, G. Dempsey, H. Surden, and P. B. Reiner, "AI Loyalty: A New Paradigm for Aligning Stakeholder Interests," *IEEE Transactions on Technology and Society*, vol. 1, pp. 128–137, 2020.
- [61] A. P. Brady and E. Neri, "Artificial Intelligence in Radiology—Ethical Considerations," *Diagnostics*, vol. 10, p. 231, Apr 2020.
- [62] W. X. Wan and T. Lindenthal, "Towards Accountability in Machine Learning Applications: A System-testing Approach," *Available at SSRN*, 2021.
- [63] G. Harrison, J. Hanson, C. Jacinto, J. Ramirez, and B. Ur, "An Empirical Study on the Perceived Fairness of Realistic, Imperfect Machine Learning Models," *Conference on Fairness, Accountability, and Transparency (Fat\* '20), January 27–30, 2020*, pp. 392–402, 2020.
- [64] S. K. McGrath and S. J. Whitty, "Accountability and responsibility defined," *International Journal of Managing Projects in Business*, vol. 11, pp. 687–707, Nov 2018.
- [65] D. Rezania, R. Baker, and A. Nixon, "Exploring project managers' accountability," *International Journal of Managing Projects in Business*, vol. 12, pp. 919–937, Nov 2019.
- [66] D. Shin and Y. J. Park, "Role of fairness, accountability, and transparency in algorithmic affordance," *Computers in Human Behavior*, vol. 98, pp. 277–284, Sep 2019.
- [67] *Artificial Intelligence Act*, R. (EU) Proposal for a regulation of the European Parliament and of the Council: Laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain union legislative acts, 2021.