Artificial Intelligence Impersonating a Human: The Impact of Design Facilitator Identity on Human Designers

Guanglu Zhang[1]
Mem. ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: glzhang@cmu.edu

Ayush Raina
Mem. ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: araina@andrew.cmu.edu

Ethan Brownell
Mem. ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: ebrownel@andrew.cmu.edu

Jonathan Cagan
Fellow ASME
Department of Mechanical Engineering,
Carnegie Mellon University,
Pittsburgh, PA 15213
e-mail: cagan@cmu.edu

Advances in artificial intelligence (AI) offer new opportunities for human–AI cooperation in engineering design. Human trust in AI is a crucial factor in ensuring an effective human–AI cooperation, and several approaches to enhance human trust in AI have been explored in prior studies. However, it remains an open question in engineering design whether human designers have more trust in an AI and achieve better joint performance when they are deceived into thinking they are working with another human designer. This research assesses the impact of design facilitator identity ("human" versus AI) on human designers through a human subjects study, where participants work with the same AI design facilitator and they can adopt their AI facilitator's design anytime during the study. Half of the participants are told that they work with an AI, and the other half of the participants are told that they work with another human participant but in fact they work with the AI design facilitator. The results demonstrate that, for this study, human designers adopt their facilitator's design less often on average when they are deceived about the identity of the AI design facilitator as another human designer. However, design facilitator identity does not have a significant impact on human designers' average performance, perceived workload, and perceived competency and helpfulness of their design facilitator in the study. These results caution against deceiving human designers about the identity of an AI design facilitator in engineering design. [DOI: 10.1115/1.4056499]

Keywords: artificial intelligence, design teams, engineering design, human–computer interaction, deception, trust, collaborative design, computer-aided design, design automation

1 Introduction

There is an emerging world of human–artificial intelligence (AI) partnership [1]. Human–AI cooperation is used in many fields to improve joint performance [2]. For example, doctors are advised by an AI to interpret medical images [3,4]; computer users employ AI prediction for the next word or phrase they want to type [5,6]. Various AIs are also applied in multiple phases of the engineering design process to solve specific design tasks alone [7–9]. Research results demonstrate that a well-trained AI can perform a specified design task as well as, or sometimes even better than, human designers [10,11]. However, when an AI advises human designers to solve a design problem, the results from a recent cognitive study show that the AI only improves the initial performance of low-performing teams but always hurts the performance of high-performing teams [12]. It is therefore crucial to understand why the AI does not improve the performance of all human designers, and more importantly, to determine how to improve the effectiveness of human–AI cooperation in the engineering design process.

Prior studies suggest that human trust in AI is a key factor in ensuring an effective human–AI cooperation [13–16]. Recent research has studied human perception and trust in AI in different fields, such as business [17], healthcare [18], robotics [19], and education [20], and various approaches have been explored to alter human trust in AI [21–24]. Among these approaches, anthropomorphism, also known as human-likeness, has been corroborated by many cognitive studies as an effective way to enhance human trust in AI [13,14,25]. Research findings of these cognitive studies show that increasing the human-likeness of an AI through anthropomorphic traits, such as human-like appearance and verbal communication, mostly improves human trust in AI and leads to better human–AI joint performance [26–33], where human trust in AI is usually measured by whether human participants are willing to accept AI advice and act on it, as well as whether human participants perceive AI as competent and helpful. Notably, participants in these prior studies are aware that they work with an AI, although the AI exhibits one or more anthropomorphic traits. It is still an open research question whether humans have more trust in AI and achieve better joint performance if humans are deceived into thinking that they are working with another human. Moreover, the identity of AI has been concealed in a few recent real-world applications, such as Google Duplex AI assistant [34], where humans are not aware that they are interacting with an AI. However, the impact of AI identity ("human" versus AI) on human practitioners has not been carefully studied yet, especially in the engineering design context.

To fill the research gap, this research assesses the impact of design facilitator identity ("human" versus AI) on the performance, behavior, and perceived workload of human designers with different design proficiency levels in the design problem-solving process through a human subjects study, where the AI design facilitator guides the human participants through designs that can be adopted anytime during the study. The human trust in AI in terms of perceived design facilitator competency and helpfulness in solving the design problem is also investigated. Notably, based on Kant's moral theory [35], lying and deception should always be prohibited, regardless of circumstances. However, consensus

[1]Corresponding author.

Contributed by the Design Theory and Methodology Committee of ASME for publication in the Journal of Mechanical Design. Manuscript received August 7, 2022; final manuscript received November 28, 2022; published online January 17, 2023. Assoc. Editor: Scarlett Miller.

has been made that Kant's absolute prohibition against lying and deception is too strict [36]. From the perspective of utilitarianism, lying and deception are morally permissible if alternative actions cannot lead to better consequences [36,37]. This research does not aim to revolve the general ethical issues related to lying and deception. Instead, the results of this research inform whether it is beneficial for human designers to be deceived about the identity of an AI as another human designer in the design problem-solving process.

In the human subjects study, each participant solves the same configuration design problem through a graphical user interface (GUI) with an AI design facilitator. Among the participants in the study, half of the participants are told that they work with an AI. The other half of the participants are told that they work with another human participant but in fact they work with the AI design facilitator. Verbal communication is not allowed in the study, but participants communicate directly with the AI design facilitator by adopting the AI facilitator's design through the GUI.

To isolate the impact of design facilitator identity ("human" versus AI) on human designers from other factors, each participant works with *the same AI design facilitator* and the AI facilitator works independently in the study. In other words, each participant observes the same design process of their AI design facilitator through the GUI in the study. Participants can adopt their AI facilitator's design anytime in the study, but the AI facilitator never adopts designs from participants during the study. In this research, the AI design facilitator is built upon a deep learning framework [38] and trained over data collected from a previous human design study that uses the same GUI and configuration design problem but does not include an AI [39]. The performance of the AI facilitator exceeds the average performance of human designers in the previous study.

This paper begins with an overview of the human subjects study in Sec. 2. The deep learning framework to construct and train the AI design facilitator is presented in Sec. 3. In Sec. 4, the results of the human subjects study are reported and the impact of design facilitator identity on human designers is analyzed. In Sec. 5, the potential reasons for the results of the human subjects study are discussed, and an argument against deceiving human designers about the identity of an AI in engineering design is provided accordingly. The paper concludes with key findings from the study and the contributions of this research.

2 Human Subjects Study Overview

Based on the protocol approved by the Carnegie Mellon University Institutional Review Board (IRB), a human subjects study is performed to assess the impact of design facilitator identity ("human" versus AI) on human designers in a configuration design problem-solving process. This section presents the GUI and the configuration design problem used in the study. The participants' information and the procedure of the study are also provided.

2.1 Graphical User Interface and Configuration Design Problem. Each participant in the human subjects study is asked to solve a truss design problem with a design facilitator through a GUI, as shown in Fig. 1. Participants can use the GUI to perform design actions, such as adding or deleting a joint, adding or deleting a two-force member between two joints, moving the position of a joint, and increasing or decreasing the thickness of a member or all members. Participants also can copy their facilitator's current design at any time in the study. In addition, the GUI indicates the factor of safety (FOS) and mass of the current truss design in real time based on the predefined loads and supports, as shown in Fig. 1 (e.g., the two down arrows represent predefined loads).

Each participant works with the same AI design facilitator and observes the same design process of their AI facilitator through the GUI in the study. Half of the participants are told that they work with an AI (No Deception Condition). The other half of the participants are told that they work with another human participant but in fact they work with the AI design facilitator (Deception Condition). Notably, the AI design facilitator works independently and never adopts participants' design in the study, but participants are not aware of such unidirectional interaction. The deep learning framework to construct the AI design facilitator and the performance of the AI design facilitator are provided in Secs. 3 and 4.1, respectively.

At the beginning of the study, the following problem statement (PS) is given to each participant:

(1) Design a bridge that spans the river, supports a load at the middle of each span, and has a factor of safety greater than 1.25.
(2) Achieve a mass that is as low as possible, preferably less than 175 kg.

Participants then start to solve the problem by designing a truss bridge from scratch.

2.2 Participants Information. Participants are recruited from an undergraduate engineering course in Carnegie Mellon University. Participants have learned the knowledge of truss design from the course before they participate in the study. Eighty-four participants are recruited and randomly assigned to two conditions. Specifically, 42 participants work with an AI design facilitator (No Deception Condition), and the other 42 participants work with a "human" facilitator (Deception Condition) in the study. Each participant receives course credit and $10 cash compensation at the end of the study. In addition, participants who achieve the top 10% performance in each condition are given an extra $10 gift card as a reward after the study. All participants are informed of the cash compensation and the extra reward at the beginning of the study. Importantly, participants who work with a "human" facilitator (Deception Condition) are given a debriefing that explains the deception about the identity of their design facilitator after the study.

2.3 Procedure of the Human Subjects Study. The human subjects study takes 40 min. Figure 2 shows the time allocation of the study. Each participant registers on entry and is given a team number and a computer number. Each participant is then assigned to one of two computer labs. Each computer lab only accommodates one condition (i.e., participants with AI facilitator or participants with "human" facilitator) at a time, and participants are not aware that the study has two conditions. Participants with AI facilitator are told that they work with an AI teammate, but they are not informed about how the AI is constructed and trained. Participants with AI facilitator also do not know the performance of the AI before they work with the AI. Participants with "human" facilitator are told that they will work with another participant in the same computer lab as their teammate, but in fact they will work with the AI design facilitator. Participants with "human" facilitator are also told that they do not know who their human teammates are in order to simulate geographically distributed teams. Notably, participants are told that they will work with an AI teammate or a human teammate at the beginning of the study, but they are not aware that their teammate never adopts their designs during the study. Although participants are told that they will work with a teammate, the teammate serves the role as a design facilitator in the study since the interaction is unidirectional; as such participants' teammate is referred to as AI design facilitator or "human" design facilitator throughout the paper.

Each participant is randomly assigned to a separate computer in the lab and is instructed to log into the GUI by entering the given team number and computer number. After logging into the GUI, each participant is given a 10-min interactive tutorial, which walks through each function of the truss design GUI. Participants are then asked to solve three truss design problems independently within 10 min to evaluate their truss design proficiency. Based on participants' truss design proficiency, the impact of design

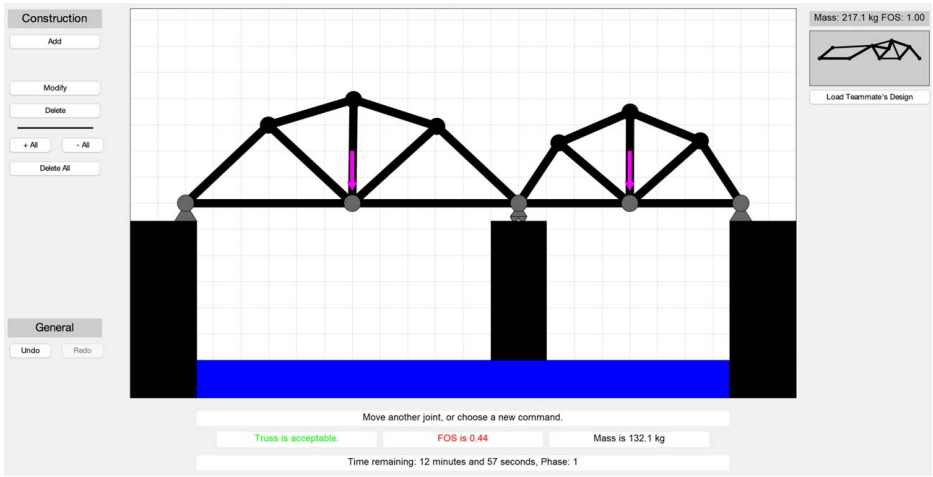

Fig. 1 Truss design GUI

facilitator identity ("human" versus AI) on human designers with different design proficiency levels is assessed as detailed in Sec. 4. These three problems, with increasing difficulty, are given in sequence to each participant, and these problems are different from the truss bridge design problem defined in Sec. 2.1. Participants are assessed an individual truss design proficiency score based on their performance in solving these three problems. The method to calculate the individual truss design proficiency score is found in the prior research [40]. The individual truss design proficiency score of each participant is only used in the post-study analysis presented in Sec. 4, and participants do not know their proficiency score in the study.

After the truss design proficiency evaluation, participants are given the PS provided in Sec. 2.1 and then design and modify truss bridges with the same AI design facilitator from scratch. As shown in Fig. 2, participants have 2 min to read the problem statement and 15 min to design truss bridges. The 15-min session includes four 3-min designs with facilitator periods. In these design with facilitator periods, participants design truss bridges on their own or copy their facilitator's design at any time onto their own design window and then modify the design. These four 3-min designs with facilitator periods are interrupted by 1-min interludes (i.e., "Select Best Design" shown in Fig. 2) to facilitate participants' interaction with the design facilitator. In each interlude, participants are asked to choose the design they will continue working on from their own current design, their own best available design, and their design facilitator's best available design.

Participants are asked to fill out a paper-pencil questionnaire after they complete the 15-min design session. The questionnaire includes eight questions. The first six questions are identical for all participants. These six questions come from the official NASA Task Load Index (NASA TLX) [41] and evaluate participants' perceived workload in the scales of mental demand, physical demand, temporal demand, performance, effort, and frustration, respectively. Participants choose a number from 0 to 100 as the answer to each question. The last two questions evaluate participants' trust in their design facilitator in terms of facilitator competency and helpfulness. The last two questions for participants with AI facilitator are "how competent is your AI teammate in solving the truss design problem?" and "how helpful was the AI teammate to you in solving the truss design problem in the study?". Similarly, the last two questions for participants with "human" facilitator are "how competent is your teammate in solving the truss design problem?" and "how helpful was your teammate to you in solving the truss design problem in the study?". Participants are asked to circle an answer from seven answer options (i.e., seven-point Likert scales) for each of the last two questions. For example, the seven answer options that participants are asked to choose from for the eighth question are "very unhelpful," "unhelpful," "somewhat unhelpful," "neither unhelpful nor helpful," "somewhat helpful," "helpful," and "very helpful."

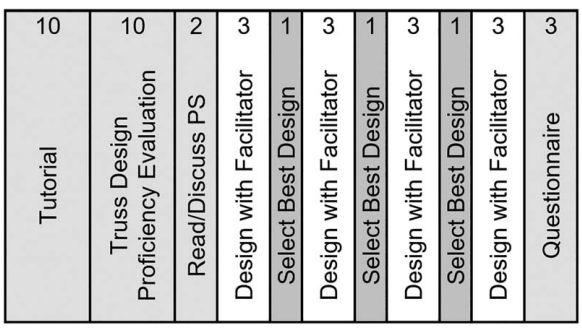

Fig. 2 Time allocation of the human subjects study (numbers indicate duration in minutes)

3 Deep Learning Framework to Construct the Artificial Intelligence Design Facilitator

The AI design facilitator employed in this study is built upon a deep learning framework introduced by Raina et al. [38]. The deep learning framework allows the AI design facilitator to perform sequential design actions and generate high-performing truss designs in the study. This section presents the deep learning framework for the AI design facilitator development. More details of the framework are found in the prior research [38].

The deep learning framework includes a design strategy network [38] and a one-step lookahead search strategy [42]. As shown in Fig. 3, the design strategy network comprises an *encoder network*, a *spatial action network*, and a *selection network*. The *encoder network* is constructed by sequential convolutional layers [43] followed by one linear layer that converts the state input of a $128 \times 128 \times 3$ pixel image to a latent representation with 512 units. The *spatial action network* then uses three linear layers to generate the parameters of the spatial region for the truss design problem. These parameters are then used to sample a set of feasible design actions in the given region of the truss design problem. The feasible design actions are then input into the *selection network* along with the latent representation of the state input derived from the *encoder network*. The *selection network* applies several linear layers to develop a combined state-action representation for every individual design action in the feasible design action set.

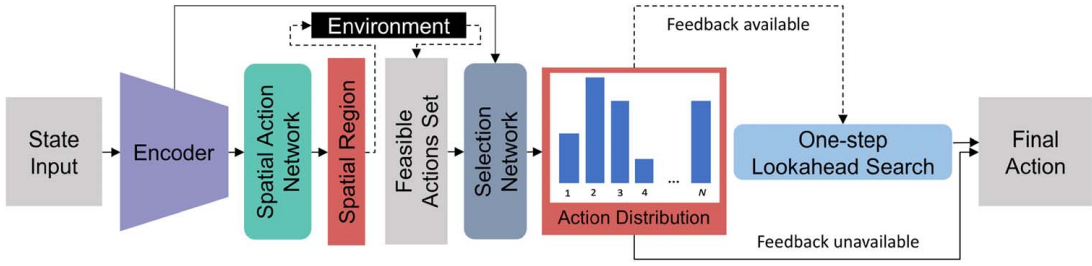

Fig. 3 Schematic representation of the deep learning framework for AI design facilitator development

The combined representation is further transformed to generate a probability distribution over the set of feasible design actions. Overall, the input to the design strategy network is an image-based truss design state, and the output is the spatial region and the probability over feasible design actions, shown as the action distribution in Fig. 3. The performance of the design strategy network is improved through a one-step lookahead search strategy. The AI design facilitator selects a design action based on the feedback from an objective function when the feedback is available (i.e., the factor of safety of the truss bridge design has a positive value) [42]. When the feedback is not available, the AI design facilitator selects the design action with the highest probability derived by the design strategy network.

The design strategy network is trained over sequential human design data collected from a previous human design study that uses the same GUI and configuration design problem but does not include an AI [39]. Each dataset is split into training, validation, and test sets using the 80-10-10 scheme. Two different loss functions are used to train the design strategy network over these datasets. Mean square error (MSE) is used to train the *encoder network* and the *spatial action network* to predict the parameters of the spatial region. Binary cross-entropy error (BCE) is used to train the *encoder network* and the *selection network* to predict the probabilities of the final design action selected by the human designers in the datasets. Details of the design strategy network training process are found in the prior research [38].

The performance of the AI design facilitator built through the trained design strategy network and the one-step lookahead search strategy over the 15-min design session in this study is provided in Sec. 4.1. Notably, the performance of the AI design facilitator exceeds the average performance of human designers in the previous study that does not include an AI. The AI facilitator performs design actions with a constant speed in this study, and the speed is set as the average speed of human designers in the previous study.

4 Results and Analysis

This section reports participants' performance, behavior, perceived workload, and perceived competency and helpfulness of their design facilitator in the human subjects study. As stated in Sec. 2.2, 84 participants are recruited and randomly assigned to two conditions (No Deception Condition and Deception Condition). The results of the participants with "human" facilitator are compared with that of the participants with AI facilitator to assess the impact of design facilitator identity ("human" versus AI) on human designers. The 42 participants with "human" facilitator and the 42 participants with AI facilitator are further split into 14 high proficiency participants, 14 middle proficiency participants, and 14 low proficiency participants, respectively, based on the truss design proficiency score each participant obtains in the 10-min truss design proficiency evaluation stated in Sec. 2.3, and the impact of design facilitator identity on human designers with different design proficiency levels is further assessed accordingly.

Notably, two-sample *t*-tests are employed to derive the *p*-values reported in Secs. 4.1 and 4.2. The *p*-values reported in Secs. 4.3 and 4.4 are computed using Mann–Whitney *U* tests since ordinal data are collected from the questionnaire at the end of the study. As a supplement to *p*-value, Cohen's *d* effect size is also reported for each comparison [44]. According to Cohen's classification [45], the *d* values of 0.2, 0.5, and 0.8 are considered as small effect size, medium effect size, and large effect size, respectively.

4.1 Performance. Participants are asked to design truss bridges that have a factor of safety greater than 1.25 and achieve a mass as low as possible, the strength-to-weight ratio (*SWR*) is therefore used to assess the performance of each participant in the study. *SWR* is defined by

$$SWR = \frac{FOS}{M} \tag{1}$$

where *M* represents the mass of a truss bridge in the unit of kilogram (kg), and *FOS* denotes the factor of safety of a truss bridge, which is dimensionless. Notably, in this study, a truss bridge design is defined to have a nonzero *SWR* value only when the *FOS* of the truss bridge is greater than 1.25 based on the design requirement defined in the problem statement. In other words, the *SWR* of a truss bridge design equals zero when the *FOS* of the truss bridge is less than or equal to 1.25.

The best available *SWR* of the AI design facilitator over the 15-min design session appears in Fig. 4. The AI design facilitator achieves the first truss bridge design that satisfies the design requirement (i.e., $FOS > 1.25$) with an *SWR* of 50.73 kg^{-1} at 2 min and 4 s and then quickly reaches the best design in the 15-min session with an *SWR* of 54.64 kg^{-1} at 2 min and 20 s. After the best design is reached, the AI design facilitator keeps modifying the truss bridge design by adding new nodes, adding new members, and increasing the thickness of each member, and the *SWR* of the following designs varies from 41.00 kg^{-1} to 54.64 kg^{-1}. Notably, since existing nodes and members are not deleted, the architecture of the truss bridge design does not change significantly after the best design is reached. The AI design facilitator works independently and never adopts participants' design. Participants can adopt the AI facilitator's current design anytime in the four 3-min designs with facilitator periods. In each of the three 1-min interludes, participants are asked to choose the design they will continue working on from their own current design, their own best available design, and the AI facilitator's best available design. Since the AI facilitator reaches the best design before the first 1-min interlude, the AI facilitator's best available designs participants can choose from in the three 1-min interludes are the same truss bridge design with the *SWR* of 54.64 kg^{-1}.

The best available *SWR* each participant achieves over the 15-min design session is tracked. As shown in Fig. 4(*a*), an average *SWR* comparison between participants with "human" facilitator and participants with AI facilitator over the 15-min design session is made to assess the impact of design facilitator identity on the average performance of human designers in the study. Figure 4(*a*) shows that the difference in average *SWR* between participants with "human" facilitator and participants with AI facilitator is not significant in the study. This result indicates that design

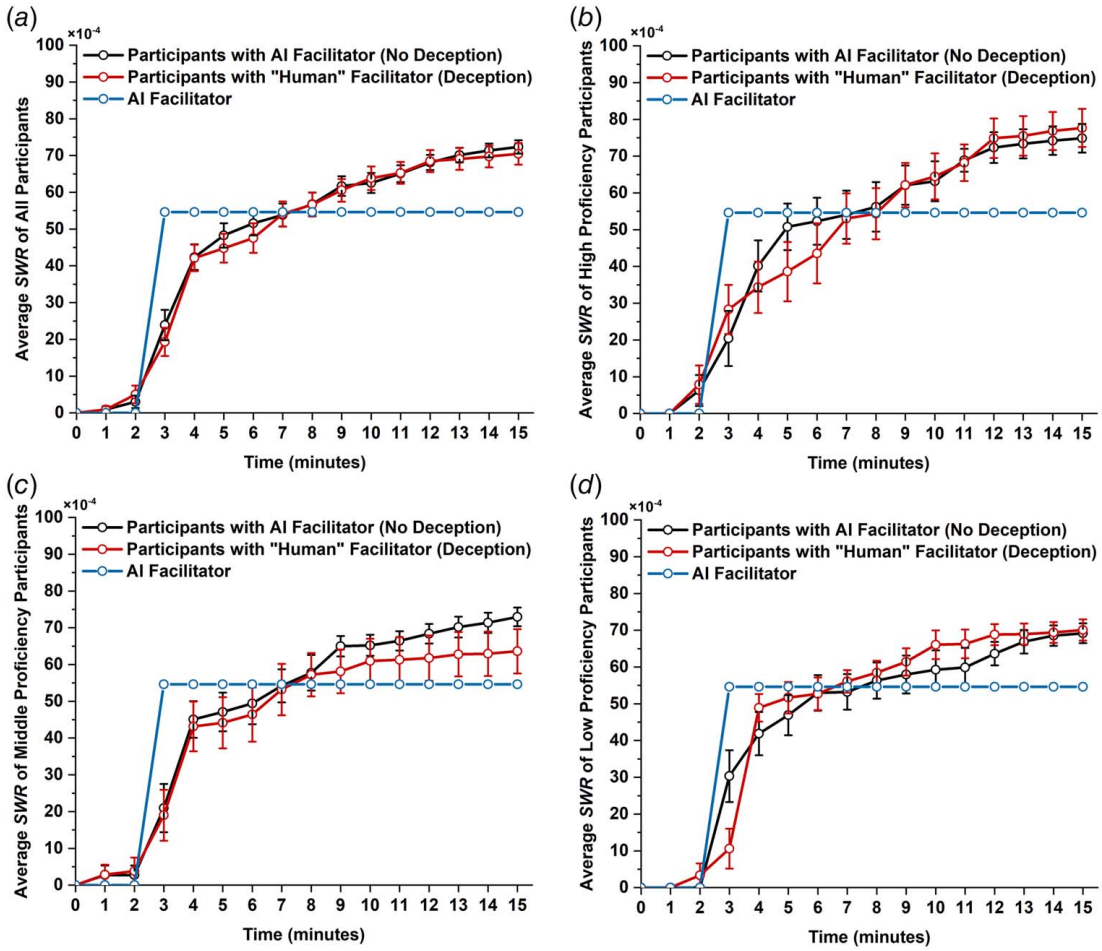

Fig. 4 *SWR* comparisons between participants with AI facilitator and participants with "human" facilitator over the 15-min design session (error bars indicate ±1 standard error): (*a*) all participants, (*b*) high proficiency participants, (c) middle proficiency participants, and (*d*) low proficiency participants

facilitator identity does not have a significant impact on the average performance of all human designers in the study.

Empirical evidence suggests that high proficiency human designers and low proficiency human designers employ different approaches to solve design problems [39,40]. To assess the impact of design facilitator identity on human designers with different design proficiency levels, high, middle, and low proficiency participants with "human" facilitator and AI facilitator are assessed and compared via average *SWR* in Figs. 4(*b*)–4(*d*), respectively.

The comparison in average *SWR* of high proficiency participants over the 15-min design session is shown in Fig. 4(*b*). The average *SWR* of high proficiency participants with "human" facilitator does not have a significant difference compared to that of high proficiency participants with AI facilitator over the study. This result indicates that design facilitator identity does not have a significant impact on the average performance of high proficiency human designers.

The comparison in average *SWR* of middle proficiency participants over the 15-min design session is shown in Fig. 4(*c*). The difference in average *SWR* at the beginning of the 15-min design session between middle proficiency participants with "human" facilitator and middle proficiency participants with AI facilitator is not significant. The average *SWR* of middle proficiency participants with "human" facilitator is 13% lower than that of middle proficiency participants with AI facilitator at the end of the 15-min design session with an effect size of 0.52 (i.e., medium effect size). However, this result does not have statistical significance ($p = 0.18$). Specifically, at the end of the 15-min design session, the average *SWR* of middle proficiency participants with "human" facilitator is $63.61 \pm 6.03\ \text{kg}^{-1}$, and the average *SWR* of middle proficiency participants with AI facilitator is $72.98 \pm 2.54\ \text{kg}^{-1}$.

The comparison in average *SWR* of low proficiency participants over the 15-min design session is shown in Fig. 4(*d*). The average *SWR* of low proficiency participants with "human" facilitator is 65% lower than that of low proficiency participants with AI facilitator before the first 1-min interlude with an effect size of 0.81 (i.e., large effect size) and a *p*-value of 0.042 (i.e., statistically significant). Specifically, at 3 min of the 15-min design session, the average *SWR* of low proficiency participants with "human" facilitator is $10.59 \pm 5.43\ \text{kg}^{-1}$, and the average *SWR* of low proficiency participants with AI facilitator is $30.34 \pm 7.05\ \text{kg}^{-1}$. Notably, only 21% (three out of 14) of low proficiency participants with "human" facilitator adopt and modify the AI facilitator's design before the first 1-min interlude. In contrast, 57% (eight out of 14) of low proficiency participants with AI facilitator adopt and modify the AI facilitator's design before the first 1-min interlude. After the first 1-min interlude, the average *SWR* of low proficiency participants with "human" facilitator does not have a significant difference compared to that of low proficiency participants with AI facilitator. These results suggest that the average performance of low proficiency human designers is significantly reduced before the first 1-min interlude in the study when these human designers are deceived into thinking they are working with another human. Design facilitator identity ("human" versus AI) does not have a significant impact on the average performance

of low proficiency human designers after the first 1-min interlude in the study.

4.2 Behavior. The number of actions and the number of adoptions of each participant in the 15-min design session are calculated. The number of participants who never adopt their facilitator's design in the 15-min design session is also counted. To assess the impact of design facilitator identity on the behavior of human designers in the study, comparisons are made between participants with "human" facilitator and participants with AI facilitator. Notably, the number of adoptions of a participant is the total number of times when the participant adopts and modifies the AI facilitator's design. The actions each participant performs before and after the participant adopts the AI facilitator's design are inspected, and an adoption of AI facilitator's design is not counted if the participant abandons the AI facilitator's design immediately after the adoption (e.g., by clicking "Undo" button or "Delete All" button in the GUI as shown in Fig. 1).

The comparison in the average number of actions between participants with "human" facilitator and participants with AI facilitator is shown in Fig. 5(*a*). Figure 5(*a*) shows that the difference in the average number of actions between participants with "human" facilitator and participants with AI facilitator is not significant. Comparisons in the average number of actions are also made between the high/middle/low proficiency participants with "human" facilitator and the corresponding high/middle/low proficiency participants with AI facilitator, respectively. As shown in Fig. 5(*a*), the average number of actions of high proficiency participants with "human" facilitator is 19% lower than that of high proficiency participants with AI facilitator with an effect size of 0.76 (i.e., medium effect size) and a p-value of 0.056 (i.e., marginally significant). Specifically, the average number of actions of high proficiency participants with "human" facilitator is 210.64 ± 16.41, and the average number of actions of high proficiency participants with AI facilitator is 260.21 ± 17.36. For low proficiency human designers, the average number of actions of low proficiency participants with "human" facilitator is 11% higher than that of low proficiency participants with AI facilitator with an effect size of 0.69 (i.e., medium effect size) and a p-value of 0.081 (i.e., marginally significant). Specifically, the average number of actions of low proficiency participants with "human" facilitator is 178.79 ± 5.04, and the average number of actions of low proficiency participants with AI facilitator is 160.79 ± 8.13. These results indicate that the average number of actions of high proficiency human designers is reduced but the average number of actions of low proficiency human designers is increased with a medium effect size and marginally significant when human designers are deceived about the identity of their AI design facilitator as another human designer.

The comparison in average number of adoptions between participants with "human" facilitator and participants with AI facilitator is shown in Fig. 5(*b*). Figure 5(*b*) shows that the average number of adoptions of participants with "human" facilitator is 50% lower than that of participants with AI facilitator with an effect size of 0.82 (i.e., large effect size) and a p-value of 3.4×10^{-4} (i.e., statistically significant). Specifically, the average number of adoptions of participants with "human" facilitator is 0.93 ± 0.09, and the average number of adoptions of participants with AI facilitator is 1.86 ± 0.23. Notably, such result holds statistical significance when the Bonferroni correction is employed to control the family-wise error rate in this study [46]. As shown in Fig. 5(*b*), the average number of adoptions of the high/middle/low proficiency participants with "human" facilitator is compared with that of the corresponding high/middle/low proficiency participants with AI facilitator, respectively. The difference in the average number of adoptions between high proficiency participants with "human" facilitator and high proficiency participants with AI facilitator is not significant ($p = 0.12$ and $d = 0.61$). In contrast, the average number of adoptions of low proficiency participants with "human" facilitator is 50% lower than that of low proficiency participants with AI facilitator with an effect size of 1.02 (i.e., large effect size) and a p-value of 0.012 (i.e., statistically significant). Specifically, the average number of adoptions of low proficiency participants with "human" facilitator is 1.07 ± 0.07, and the average number of adoptions of low proficiency participants with AI facilitator is 2.14 ± 0.38. These results suggest that the average number of adoptions of human designers, especially low proficiency human designers, is significantly reduced in the study with a large effect size when human designers are deceived into thinking they are working with another human designer.

The comparisons in the number of participants who never adopt their facilitator's design between participants with "human" facilitator and participants with AI facilitator with different design proficiency levels are shown in Fig. 5(*c*). Notably, as shown in Fig. 5(*c*), 29% (four out of 14) of high proficiency participants with "human" facilitator never adopt their facilitator's design. In contrast, 43% (six out of 14) of high proficiency participants with AI facilitator never adopt their facilitator's design. This result shows that more high proficiency human designers adopt and modify their facilitator's design at least one time in the study when they are told that they work with another human designer rather than an AI.

4.3 Perceived Workload. Participants are asked to fill out a paper-pencil questionnaire after they complete the 15-min design session. The first six questions in the questionnaire come from the official NASA Task Load Index (NASA TLX) [41] and evaluate participants' perceived workload on the scales of mental demand, physical demand, temporal demand, performance, effort, and frustration, respectively. Participants choose a number from 0 to 100 as the answer to each question. The answers from participants with "human" facilitator are compared to the answers from participants with AI facilitator to assess the impact of design facilitator identity on the perceived workload of human designers in the study.

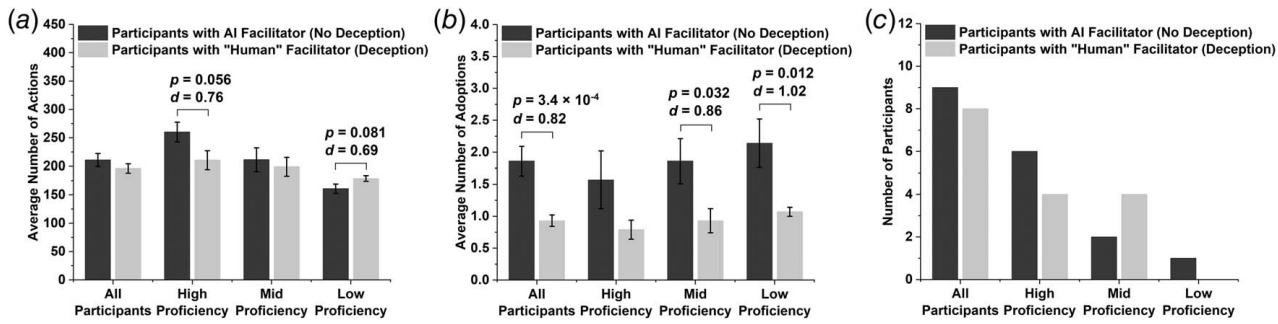

Fig. 5 Behavior comparisons between participants with AI facilitator and participants with "human" facilitator over the 15-min design session (error bars indicate ±1 standard error): (*a*) average number of actions, (*b*) average number of adoptions, and (*c*) number of participants who never adopt facilitator's design

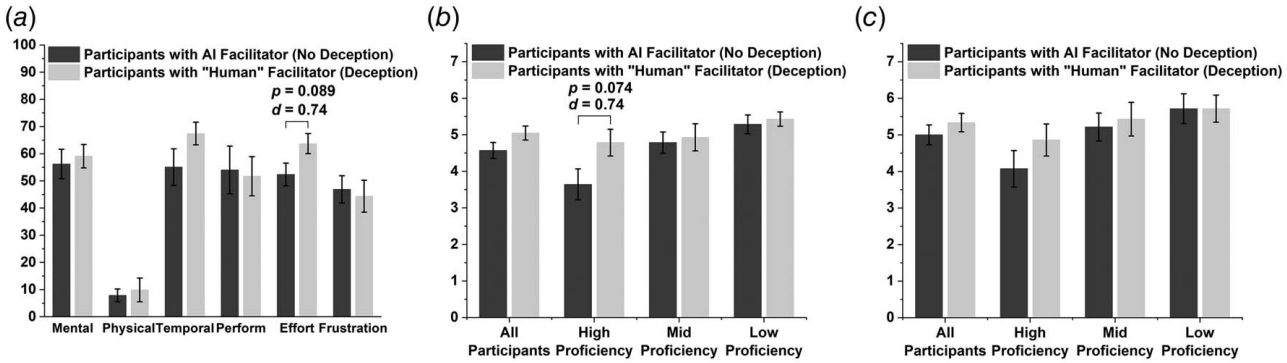

Fig. 6 Comparisons in perceived workload and trust in facilitator between participants with AI facilitator and participants with "human" facilitator in the questionnaire (error bars indicate ±1 standard error): (*a*) NASA TLX of low proficiency human designers, (*b*) perceived facilitator competency, and (*c*) perceived facilitator helpfulness

Comparisons in the answers of NASA TLX are first made between participants with "human" facilitator and participants with AI facilitator. No significant result is found in these comparisons, which suggests that design facilitator identity does not have a significant impact on the perceived workload of all human designers in the study. The answers of NASA TLX from the high/middle/low proficiency participants with "human" facilitator are then compared to that from the corresponding high/middle/low proficiency participants with AI facilitator, respectively. The only marginally significant result is found in the comparison of the perceived effort between low proficiency participants with "human" facilitator and low proficiency participants with AI facilitator, and the results of all other comparisons are not significant ($p > 0.1$). As shown in Fig. 6(*a*), low proficiency participants with "human" facilitator believe that they spend more effort on average to accomplish their level of performance compared to low proficiency participants with AI facilitator with an effect size of 0.74 (i.e., medium effect size) and a p-value of 0.089 (i.e., marginally significant). This result shows that the perceived effort of low proficiency human designers is raised in the study with a medium effect size and marginally significant when these human designers are deceived about the identity of their AI design facilitator as another human designer.

4.4 Perceived Competency and Helpfulness of Design Facilitator. As stated in Sec. 2.3, the last two questions in the questionnaire evaluate each participant's perceived competency and helpfulness of their design facilitator in the study. Participants are asked to circle an answer from seven answer options (i.e., seven-point Likert scale) for each of the last two questions. The answers to the last two questions from participants with "human" facilitator are compared with the corresponding answers from participants with AI facilitator to assess the impact of design facilitator identity on human designers' perceived competency and helpfulness of their design facilitator. Notably, in the seven-point Likert scale, one represents the most negative answer (e.g., "very unhelpful" for the eighth question), and seven represents the most positive answer (e.g., "very helpful" for the eighth question).

The comparisons in the answers about design facilitator competency between participants with "human" facilitator and participants with AI facilitator with different design proficiency levels are shown in Fig. 6(*b*). The only marginally significant result is found in the comparison between high proficiency participants with "human" facilitator and high proficiency participants with AI facilitator, and the other three results are not significant ($p > 0.1$). High proficiency participants with "human" facilitator believe that their design facilitator is more competent in solving the truss design problem compared to high proficiency participants with AI facilitator with an effect size of 0.74 (i.e., medium effect size) and a p-value of 0.074 (i.e., marginally significant). This result suggests that high proficiency human designers' perceived competency of their design facilitator is improved with a medium effect size and marginally significant when they are deceived about the identity of their AI design facilitator as another human designer.

The comparisons in the answers about design facilitator helpfulness between participants with "human" facilitator and participants with AI facilitator with different design proficiency levels are shown in Fig. 6(*c*). No significant result is found in these comparisons ($p > 0.1$), which suggests that design facilitator identity does not have a significant impact on human designers' perceived helpfulness of their design facilitator in the study.

5 Discussion

The results of the human subjects study presented in Sec. 4 suggest that the average performance of low proficiency human designers is reduced before the first interlude when they are deceived into thinking they are working with another human designer, but otherwise design facilitator identity ("human" versus AI) does not have a significant impact on the average performance of high or low proficiency human designers in the study. Three potential reasons behind these results are discussed below.

First, low proficiency human designers are more willing to adopt designs from an AI facilitator rather than another human designer at the beginning of the study. Although low proficiency participants with "human" facilitator do not report a lower level of trust in their design facilitator in the post-study questionnaire compared to low proficiency participants with AI facilitator, only 21% (three out of 14) of low proficiency participants with "human" facilitator adopt and modify their facilitator's design before the first interlude. In contrast, before the first interlude, 57% (eight out of 14) of low proficiency participants with AI facilitator adopt and modify their facilitator's design. Importantly, low proficiency human designers struggle to create an acceptable truss bridge design at the beginning of the study. Among low proficiency participants, only the low proficiency participants who adopt designs from their AI facilitator are able to create a truss bridge design that satisfies the design requirement before the first interlude. Less number of low proficiency human designers adopts their facilitator's design before the first interlude when they are deceived about the identity of their AI facilitator as another human and therefore the average performance of low proficiency human designers is reduced at the beginning of the study.

Second, the three interludes greatly promote low proficiency human designers to adopt their facilitator's designs in the study. In each interlude, participants are asked to choose the design they will continue working on from their own current design, their own best available design, and their facilitator's best available design. The first interlude raises the percentage of low proficiency participants with "human" facilitator who adopt and modify their facilitator's design from 21% (three out of 14) to 93% (13 out of 14), and all low proficiency participants with "human" facilitator

adopt and modify their facilitator's design at least one time by the end of the study. As stated in Sec. 4.1, the AI design facilitator reaches its best truss bridge design before the first interlude, and the architecture of the truss bridge the AI design facilitator creates does not change significantly after the best design is reached; low proficiency human designers therefore may not be able to earn extra benefit from adopting and modifying their AI facilitator's design multiple times after the first interlude in the study. Thus, design facilitator identity does not have a significant impact on the average performance of low proficiency human designers after the first interlude, although the average number of adoptions of low proficiency human designers is significantly reduced in the study when they are deceived about the identity of their AI design facilitator as another human designer.

Third, high proficiency human designers trust another human designer more than an AI design facilitator, and high proficiency human designers become reliant when they are told that they work with another human designer rather than an AI. In the post-study questionnaire, high proficiency participants with "human" facilitator believe that their design facilitator is more competent in solving the truss design problem compared to high proficiency participants with AI facilitator. Only 29% (four out of 14) of high proficiency participants with "human" facilitator never adopt their facilitator's design in the study. In contrast, 43% (six out of 14) of high proficiency participants with AI facilitator never adopt their facilitator's design in the study. However, the average number of actions of high proficiency participants with "human" facilitator is 19% lower than that of high proficiency participants with AI facilitator. As suggested by these results, high proficiency human designers' trust in their design facilitator is enhanced when they are deceived into thinking they are working with another human designer rather than an AI, and thus more high proficiency participants with "human" facilitator take advantage of the truss bridge design created by their AI design facilitator in the study. However, with the enhanced trust in their design facilitator, high proficiency participants with "human" facilitator may rely on their design facilitator and become less motivated to keep creating better truss bridge designs in the study. The advantage high proficiency participants with "human" facilitator take from their design facilitator may be offset by their own uncertainty and reliance on their design facilitator in the study, and design facilitator identity ("human" versus AI) therefore does not have a significant impact on the average performance of high proficiency human designers in the study.

Based on the results of the study and the potential reasons discussed above, we caution against deceiving human designers about the identity of an AI design facilitator in engineering design. The deception about AI identity as another human is different from the anthropomorphic traits employed in prior studies, such as human-like appearance and verbal communication, and such deception does not improve human trust in AI and human–AI joint performance for all human designers with different design proficiency levels in this study. Specifically, for high proficiency human designers, in this study, although their perceived design facilitator competency is improved when they are deceived about the identity of their AI design facilitator as another human designer, such deception does not raise the average number of adoptions and does not boost their average performance. For low proficiency human designers, the deception about the identity of an AI design facilitator does not benefit them at all in the study. Rather, such deception reduces the average number of adoptions and hurts the average performance of low proficiency human designers at the beginning of the study. Thus, besides ethical concerns [37], practitioners need to consider potential detrimental effects to human designers, especially to low proficiency human designers, when they plan to deceive human designers about the identity of their AI design facilitator in solving engineering design problems.

This study has several limitations that offer opportunities for future research. First, the human subjects study presented in this paper employs a typical configuration design problem. To generalize the findings from this study outside of engineering design, future research could explore the impact of AI identity ("human" versus AI) on human–AI cooperation using various tasks in other fields, such as robotic control and AI-assisted vehicle driving, where AI keeps giving advice in the human decision-making process. In addition, the AI design facilitator works independently and never adopts human participants' designs in this study. In other words, the interaction between human designers and the AI design facilitator is unidirectional in this study. Future research could create AI teammates that support bidirectional human–AI interaction. For example, in an engineering design process, both human designers and their AI teammates would be able to adopt and modify each other's designs, which also allows for the study of bidirectional trust between human and AI (i.e., human trust in AI and AI trust in human).

6 Conclusions

The impact of design facilitator identity ("human" versus AI) is assessed through a human subjects study. All participants work with the same AI design facilitator in the study. Half of the participants are told that they work with an AI, and the other half of the participants are told that they work with another human participant but in fact they work with the AI design facilitator. The results of the study indicate that human designers adopt their facilitator's design less often on average when they are deceived into thinking they are working with another human designer rather than an AI, but design facilitator identity does not have a significant impact on human designers' average performance, perceived workload, and perceived competency and helpfulness of their design facilitator in the study.

This research cautions against deceiving human designers about the identity of their AI design facilitator in engineering design since such deception is different from the anthropomorphic traits employed in prior studies (e.g., human-like appearance and verbal communication) and may not be able to benefit human designers, especially low proficiency human designers. This research also suggests that high proficiency human designers and low proficiency human designers may have different levels of perceived competency and helpfulness of an AI, which is a factor that needs to be considered in future cognitive studies and applications involving the use of AI in design.

Acknowledgment

This material is partially supported by the Air Force Office of Scientific Research through Grant FA9550-18-0088. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the sponsor.

Conflict of Interest

There are no conflicts of interest.

Data Availability Statement

The datasets generated and supporting the findings of this article are obtainable from the corresponding author upon reasonable request.

References

[1] National Science and Technology Council, 2016, *Preparing for the Future of Artificial Intelligence*, National Science and Technology Council Report, Washington, DC, https://obamawhitehouse.archives.gov/sites/default/files/whitehouse_files/microsites/ostp/NSTC/preparing_for_the_future_of_ai.pdf

[2] Wilson, H. J., and Daugherty, P. R., 2018, "Collaborative Intelligence: Humans and AI Are Joining Forces," Harvard Business Rev., **96**(4), pp. 114–123. https://hbr.org/2018/07/collaborative-intelligence-humans-and-ai-are-joining-forces

[3] The Chartered Institute of Ergonomics and Human Factors, 2021, *Human Factors and Ergonomics in Healthcare AI*, The Chartered Institute of Ergonomics and Human Factors (CIEHF) White Paper, Wootton Park, UK, https://ergonomics.org.uk/resource/human-factors-in-healthcare-ai.html
[4] Razzak, M. I., Naz, S., and Zaib, A., 2018, "Deep Learning for Medical Image Processing: Overview, Challenges and the Future," *Classification in BioApps*, N. Dey, A. S. Ashour, and S. Borra, eds., Springer International Publishing AG, Gewerbestrasse, Switzerland, pp. 323–350.
[5] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., and Dean, J., 2013, "Distributed Representations of Words and Phrases and Their Compositionality," Proceedings of the 26th International Conference on Neural Information Processing Systems, Vol. 2, Lake Tahoe, NV, Dec. 5–8.
[6] Manning, C., and Schutze, H., 1999, *Foundations of Statistical Natural Language Processing*, MIT Press, Cambridge, MA.
[7] Chen, H. Q., Honda, T., and Yang, M. C., 2013, "Approaches for Identifying Consumer Preferences for the Design of Technology Products: A Case Study of Residential Solar Panels," ASME J. Mech. Des., **135**(6), p. 061007.
[8] Camburn, B., He, Y., Raviselvam, S., Luo, J., and Wood, K., 2020, "Machine Learning-Based Design Concept Evaluation," ASME J. Mech. Des., **142**(3), p. 031113.
[9] Williams, G., Meisel, N. A., Simpson, T. W., and McComb, C., 2019, "Design Repository Effectiveness for 3D Convolutional Neural Networks: Application to Additive Manufacturing," ASME J. Mech. Des., **141**(11), p. 111701.
[10] Lopez, C. E., Miller, S. R., and Tucker, C. S., 2019, "Exploring Biases Between Human and Machine Generated Designs," ASME J. Mech. Des., **141**(2), p. 021104.
[11] Raina, A., McComb, C., and Cagan, J., 2019, "Learning to Design From Humans: Imitating Human Designers Through Deep Learning," ASME J. Mech. Des., **141**(11), p. 111102.
[12] Zhang, G., Raina, A., Cagan, J., and McComb, C., 2021, "A Cautionary Tale About the Impact of AI on Human Design Teams," Des. Stud., **72**, p. 100990.
[13] Glikson, E., and Woolley, A. W., 2020, "Human Trust in Artificial Intelligence: Review of Empirical Research," Acad. Manage. Ann., **14**(2), pp. 627–660.
[14] Siau, K., and Wang, W., 2018, "Building Trust in Artificial Intelligence, Machine Learning, and Robotics," Cutter Business Technol. J., **31**(2), pp. 47–53. https://www.cutter.com/article/building-trust-artificial-intelligence-machine-learning-and-robotics-498981.
[15] Hoff, K. A., and Bashir, M., 2015, "Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust," Hum. Factors, **57**(3), pp. 407–434.
[16] Lee, J. D., and See, K. A., 2004, "Trust in Automation: Designing for Appropriate Reliance," Hum. Factors, **46**(1), pp. 50–80.
[17] Jan, S. T., Ishakian, V., and Muthusamy, V., 2020, "AI Trust in Business Processes: The Need for Process-Aware Explanations," Proceedings of the AAAI Conference on Artificial Intelligence, New York, Feb. 7–12, pp. 13403–13404.
[18] Asan, O., Bayrak, A. E., and Choudhury, A., 2020, "Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians," J. Med. Internet Res., **22**(6), p. e15154.
[19] Wang, W., and Siau, K., 2019, "Artificial Intelligence, Machine Learning, Automation, Robotics, Future of Work and Future of Humanity: A Review and Research Agenda," J. Database Manage., **30**(1), pp. 61–79.
[20] Nazaretsky, T., Ariely, M., Cukurova, M., and Alexandron, G., 2022, "Teachers' Trust in AI-Powered Educational Technology and a Professional Development Program to Improve It," Brit. J. Educ. Technol., **53**(4), pp. 914–931.
[21] Wang, W., and Benbasat, I., 2007, "Recommendation Agents for Electronic Commerce: Effects of Explanation Facilities on Trusting Beliefs," J. Manage. Inform. Syst., **23**(4), pp. 217–246.
[22] Pieters, W., 2011, "Explanation and Trust: What to Tell the User in Security and AI?," Ethics Inform. Technol., **13**(1), pp. 53–64.
[23] Chong, L., Zhang, G., Goucher-Lambert, K., Kotovsky, K., and Cagan, J., 2022, "Human Confidence in Artificial Intelligence and in Themselves: The Evolution and Impact of Confidence on Adoption of AI Advice," Comput. Hum. Behav., **127**, p. 107018.
[24] Gillath, O., Ai, T., Branicky, M. S., Keshmiri, S., Davison, R. B., and Spaulding, R., 2021, "Attachment and Trust in Artificial Intelligence," Comput. Hum. Behav., **115**, p. 106607.
[25] Li, M., and Suh, A., 2022, "Anthropomorphism in AI-Enabled Technology: A Literature Review," Electron. Markets, pp. 1–31.
[26] de Visser, E. J., Krueger, F., McKnight, P., Scheid, S., Smith, M., Chalk, S., and Parasuraman, R., 2012, "The World Is Not Enough: Trust in Cognitive Agents," Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, Boston, MA,, Oct. 22–26, pp. 263–267.
[27] Pak, R., Fink, N., Price, M., Bass, B., and Sturre, L., 2012, "Decision Support Aids With Anthropomorphic Characteristics Influence Trust and Performance in Younger and Older Adults," Ergonomics, **55**(9), pp. 1059–1072.
[28] Kulms, P., and Kopp, S., 2019, "More Human-Likeness, More Trust? The Effect of Anthropomorphism on Self-Reported and Behavioral Trust in Continued and Interdependent Human-Agent Cooperation," Proceedings of Mensch Und Computer 2019, Hamburg, Germany, Sept. 8–11, pp. 31–42.
[29] de Visser, E. J., Monfort, S. S., Goodyear, K., Lu, L., O'Hara, M., Lee, M. R., Parasuraman, R., and Krueger, F., 2017, "A Little Anthropomorphism Goes a Long Way: Effects of Oxytocin on Trust, Compliance, and Team Performance With Automated Agents," Hum. Factors, **59**(1), pp. 116–133.
[30] de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., and Parasuraman, R., 2016, "Almost Human: Anthropomorphism Increases Trust Resilience in Cognitive Agents," J. Exp. Psychol.: Appl., **22**(3), pp. 331–349.
[31] Verberne, F. M., Ham, J., and Midden, C. J., 2015, "Trusting a Virtual Driver That Looks, Acts, and Thinks Like You," Hum. Factors, **57**(5), pp. 895–909.
[32] Pelau, C., Dabija, D.-C., and Ene, I., 2021, "What Makes an AI Device Human-Like? The Role of Interaction Quality, Empathy, and Perceived Psychological Anthropomorphic Characteristics in the Acceptance of Artificial Intelligence in the Service Industry," Comput. Hum. Behav., **122**, p. 106855.
[33] Fox, J., Ahn, S. J., Janssen, J. H., Yeykelis, L., Segovia, K. Y., and Bailenson, J. N., 2015, "Avatars Versus Agents: A Meta-Analysis Quantifying the Effect of Agency on Social Influence," Hum.-Comput. Interact., **30**(5), pp. 401–432.
[34] O'Leary, D. E., 2019, "GOOGLE'S Duplex: Pretending to Be Human," Intell. Syst. Account. Finance Manage., **26**(1), pp. 46–53.
[35] Kant, I., 1948, *Moral Law: Groundwork of the Metaphysics of Morals*, Routledge, New York.
[36] Carson, T. L., 2010, *Lying and Deception: Theory and Practice*, Oxford University Press, Oxford, UK.
[37] Shim, J., and Arkin, R. C., 2013, "A Taxonomy of Robot Deception and Its Benefits in HRI," Proceedings of the 2013 IEEE International Conference on Systems, Man, and Cybernetics, Manchester, UK, Oct. 13–16,pp. 2328–2335.
[38] Raina, A., Cagan, J., and McComb, C., 2022, "Design Strategy Network: A Deep Hierarchical Framework to Represent Generative Design Strategies in Complex Action Spaces," ASME J. Mech. Des., **144**(2), p. 021404.
[39] McComb, C., Cagan, J., and Kotovsky, K., 2015, "Rolling With the Punches: An Examination of Team Performance in a Design Task Subject to Drastic Changes," Des. Stud., **36**, pp. 99–121.
[40] Brownell, E., Cagan, J., and Kotovsky, K., 2021, "Only As Strong As the Strongest Link: The Relative Contribution of Individual Team Member Proficiency in Configuration Design," ASME J. Mech. Des., **143**(8), p. 081402.
[41] Hart, S. G., and Staveland, L. E., 1988, "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research," Adv. Psychol., **52**, pp. 139–183.
[42] Raina, A., Puentes, L., Cagan, J., and McComb, C., 2021, "Goal-Directed Design Agents: Integrating Visual Imitation With One-Step Lookahead Optimization for Generative Design," ASME J. Mech. Des., **143**(12), p. 124501.
[43] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P., 1998, "Gradient-Based Learning Applied to Document Recognition," Proc. IEEE, **86**(11), pp. 2278–2324.
[44] Sullivan, G. M., and Feinn, R., 2012, "Using Effect Size—Or Why the P Value Is Not Enough," J. Graduate Med. Educ., **4**(3), pp. 279–282.
[45] Cohen, J., 1988, *Statistical Power Analysis for the Behavioral Sciences*, Lawrence Erlbaum Associates, Inc., New York.
[46] Miller, R. G., 1981, *Simultaneous Statistical Inference*, Springer-Verlag, New York.