Dissociation Between Users' Explicit and Implicit Attitudes Toward Artificial Intelligence: An Experimental Study

Valentina Fietta, Francesca Zecchinato, Brigida Di Stasi, Mirko Polato, *Associate Member, IEEE*, and Merylin Monaro

***Abstract*—The latest developments in the field of artificial intelligence (AI) have given rise to many ethical and socio-economic concerns. Nonetheless, the impact of AI technologies is evident and tangible in our everyday life. This dichotomy leads to mixed feelings toward AI: people recognize the positive impact of AI, but they also show concerns, especially about their privacy and security. In this article, we try to understand whether the implicit and explicit attitudes toward AI are coherent. We investigated explicit and implicit attitudes toward AI by combining a self-report measure and an implicit measure, i.e., the implicit association test. We analyzed the explicit and implicit responses of 829 participants. Results revealed that while most of the participants explicitly express a positive attitude toward AI, their implicit responses seem to point in the opposite direction. Results also show that, in both the explicit and implicit measures, females show a more negative attitude than males, and people who work in the field of AI are inclined to be positive toward AI.**

***Index Terms*—Attitudes toward artificial intelligence (AI), explicit–implicit cognition, implicit association test (IAT), implicit attitudes.**

I. Introduction

BROADLY speaking, artificial intelligence (AI) refers to machine systems that are capable of sophisticated (i.e., intelligent) information processing [1]. Nowadays, AI systems are increasingly part of our daily lives, affecting our society and economy in ways we are often not conscious about [2]. AI systems are widely present as piece of software in computers, mobile phones, and TV applications, to name a few. For example, AI can predict our preferences and future purchases and make suggestions based on our previous choices. AI influences our existence substantially, and further advances in the field of AI have the potential to impact nearly all aspects of society: the job market, transportation, healthcare, education, and national security [1]. Although AI frequently enters the political and technological debates, public opinion has not yet faced discussions on the benefits and risks of using AI in the long-term [1]. Indeed, the assessment of the public's attitudes toward AI has received little attention and, to date, only a few scientific studies have systematically investigated opinions about AI and accounted for those individual factors that implicitly mediate attitudes and user experience toward AI systems and devices [3], [4]. Extensive knowledge of attitudes and opinions toward AI may increase the acceptance of technology [5], and this is especially relevant in some areas, such as medicine and public health, where AI can bring significant improvements. Currently, it is still unclear whether the worldwide population does have a full understanding of what AI technologies really are and their practical applications [1]. The public opinion concerning AI appears to be mixed: on one hand, people recognize the positive impacts that AI technologies have on their lives and are optimistic about the future use; on the other hand, significant concerns emerge, especially in relation to job loss, privacy, and excess of control of AI systems over humans [6]–[8]. Moreover, AI technologies are developing rapidly and people's opinions are influenced by contextual variables [5].

This study, by combining self-report measures and the implicit association test (IAT), aims to investigate people's implicit and explicit attitudes toward AI systems, as compared to human-based activities and services. In line with previous studies and with the evidence that individuals tend to show an implicit preference for what is perceived as similar, as opposed to dissimilar [9], we hypothesize that participants will show more positive attitudes and more trust toward human and human-controlled services, as compared to AI technologies, regardless of their explicitly declared (i.e., self-reported) opinions. Moreover, previous studies [10] demonstrated that demographic characteristics and familiarity with AI play an important role in the attitude toward these technologies. Thus, in this article it will be investigated how implicit and explicit attitudes toward AI change according to demographic variables (gender, age, education level) and familiarity with the AI field.

Crucially, a systematic evaluation of the level of acceptance, opinions and preferences toward AI by different people and in different sectors, adopting both implicit and explicit measures, can provide a reliable guide to further develop and improve AI

Manuscript received March 12, 2021; revised June 5, 2021, July 29, 2021, and September 21, 2021; accepted October 24, 2021. This article was recommended by Associate Editor Emilia I. Barakova. *(Corresponding author: Merylin Monaro.)*

Valentina Fietta, Francesca Zecchinato, Brigida Di Stasi, and Merylin Monaro are with the Department of General Psychology, University of Padova, 35131 Padova, Italy (e-mail: valefietta.vf@gmail.com; francesca.zecchinato@studenti.unipd.it; brigida.distasi@studenti.unipd.it; merylin.monaro@unipd.it).

Mirko Polato is with the Department of Computer Science, University of Turin, 10149 Turin, Italy (e-mail: mpolato@math.unipd.it).

Color versions of one or more figures in this article are available at https://doi.org/10.1109/THMS.2021.3125280.

Digital Object Identifier 10.1109/THMS.2021.3125280

technologies and, most importantly, to ameliorate user experience in relation to AI.

II. Related Works

A. Previous Studies on Explicit Attitudes Toward AI

To date, attitudes toward AI and the level of acceptance of intelligent technologies have been just marginally investigated by the social sciences. Moreover, data have been collected mostly through explicit techniques, like questionnaires and structured interviews.

In 2019, Zhang and Dafoe conducted an extensive survey addressing the American public's attitudes toward AI and AI governance [10]. The questions included topics such as workplace automation, attitudes regarding international cooperation, the trust of the subjects in AI and its impact in governance. The responses of US people to the survey revealed that many of them looked positively at AI technologies. In particular, the greatest support came from wealthy, educated, male individuals, and from those who had direct experience with AI technologies. However, 82% of the surveyed thought that AI and robots should be carefully managed, due to the possible AI-enhanced cyber-attacks, surveillance, and global manipulation. Crucially, the survey highlighted that a majority of Americans believe that high-level machine intelligence would be harmful to humans [10].

The Special Eurobarometer 460, a survey among all 28 EU countries, detected the attitudes of subjects toward the impact of digitalization and automation in daily life [10]. The results indicated that 75% of the Europeans believed it has a positive impact on the economy, 67% on the quality of life, and 64% on society. Again, differences emerged among different demographic groups: men, younger respondents, well-educated, daily Internet users and those with less financial stressors were more positive toward digitalization and robotization. Although having a generally positive view of AI, 72% of the participants believed that robots and AI steal people's jobs and 88% of the respondents agreed that robots and AI are technologies that require careful management. Further, it was found that a greater knowledge of AI increases the positive attitude toward it [10]. Similarly, another study revealed that people who had previous experience with robots showed more positive attitudes when interacting with them than those who had not [11]. It was also found that the cultural background had a significant effect on participants' attitudes toward robots.

In a study investigating opinions of 352 machine learning experts of different nationalities, Grace *et al.* [7] found that this population predicted that AI would outperform humans in many activities over the next 10 years (such as translating languages, writing high-school essays, driving trucks, working as a surgeon etc.) and that the advances in AI would lead to the automatization of all human jobs in 120 years. A brief questionnaire aimed at assessing experts' opinions related to the future progress in AI revealed that experts expect a 50% chance that higher-level intelligence will be developed by 2040–2050 and 90% chance by 2075 [12]. According to the responders, there would be a 1/3 chance that superintelligence will have a "bad" or "extremely bad" impact for humans. One of the most complex issues concerns the perception of moral responsibility. Some authors suggested that in high stake scenarios people might attribute to AI the same level of causal responsibility and blame as they attribute to humans, and they might expect that both humans and AI can justify their decisions in the same way [13]. Differently, other studies demonstrated that people may not be willing to blame machines for moral harm [14].

Finally, there is evidence suggesting that the current global health crisis may have increased people's acceptance of AI technologies. For instance, recent empirical studies conducted in the tourism industry showed that, in the presence of the COVID-19 threat, consumers preferred robot-staffed hotels rather than human-staffed ones. Notably, these findings are in contrast with many of the studies conducted before COVID-19, which indicated a preference for human service rather than robot service in hotels and may be mainly due to the health crisis [5].

In the final analysis, the current literature suggests that people have a positive explicit attitude toward AI; however, at the same time, they show concerns about the possible consequence of applying AI for certain purposes and prospect future scenarios, where AI technology replaces humans and becomes potentially dangerous for them. Differences in explicit attitudes toward AI emerged according to gender, age, education, and previous experience/work in the field of AI. Attitudes are also influenced by external and cultural factors, which can lead to a rapid change in public opinion.

B. Implicit Attitudes and IAT

Our cognition can operate in explicit and implicit modes, with both modes having the potential to influence our behavior and judgments, although acting through different processes and influenced by different sources of information [15]–[19]. Explicit cognition refers to deliberate processes that require the expenditure of mental effort and are under the individual's conscious control, whereas implicit cognition encompasses more automatic processes that are likely to operate without awareness [16], [17], [20]. Attitudes are also subject to these two mechanisms. They are defined as memory associations between a target object and an evaluation of that object [21]. The stronger the association between the object and its evaluation, the greater the possibility that the evaluation is spontaneously activated in front of the object. Evaluations that come to mind very quickly are called automatic or implicit attitudes, while those needing reflections are referred to as deliberative or explicit attitudes. Consequently, people are mostly aware of their explicit evaluations, but they are not necessarily aware of their implicit evaluations. Previous studies demonstrated that implicit and explicit attitudes do not always coincide: the explicit attitude toward a target object can be negative whereas the implicit attitude is positive or vice versa [22], [23]. This phenomenon is called implicit–explicit discrepancy (IED), or implicit ambivalence toward the attitude object. In these circumstances, the subject is aware of his positive but not of his negative reaction to the object (or vice versa) or is aware of having both positive and negative reactions but denies that one reaction is valid or believes it derives from an external source (e.g., from social media).

Different models have been proposed to explain IED [24], [25]. The Associative Propositional Evaluation Model (APE)

assumes that there are two different systems by which attitudes are formed: 1) a slow system, responsible for implicit attitudes, which are shaped through the repeated pairings between an object and the related evaluations; 2) a fast system, responsible for forming explicit attitudes, which operates by cognitive processes of higher level and is affected by explicit processing goals [24]. Thus, explicit attitudes can change quickly in response to new information, but the old implicit attitude can also remain in memory and influence subsequent behaviors [26], [27]. According to the meta-cognitive model, this results in dissonance; indeed, old, and new attitudes can interact producing evaluative responses consistent with a state of implicit ambivalence [35]. Prior research has also demonstrated that the magnitude of IED (large vs. small IED) plays an important role in processing new information related to the target object: the greater the discrepancy between implicit and explicit attitude, the greater the scrutiny of information [28]. People with a low explicit and a high implicit prejudice toward a target object, or vice versa, seem to process new information carefully with the attempt to reduce the discomfort deriving from ambivalence and dissonance [28].

The differences between explicit and implicit cognitions implicate that their measurement requires different methods and tools. Indeed, while explicit cognition, presuming accurate introspection, can be assessed using direct and self-report measures (e.g., questionnaires), the investigation of implicit cognitions requires indirect measures, which allow the concealment of what is being assessed and the elicitation of an automatic (i.e., not consciously controlled) response from the examinee [16]. This characteristic makes implicit measures particularly important, providing the opportunity to overcome key limitations of self-report measures, such as social desirability biases or lack of awareness [16].

A prominent and widespread tool for the measure of implicit cognitions is represented by the IAT, a tool developed and validated to detect the strength of the automatic association between concepts [29], [30]. The classical IAT consists in a computer-based task in which participants are asked to classify visual stimuli (e.g., words or images) using two keyboard keys, while their reaction times are recorded [29]. The instrument builds on the assumption that it is easier, hence faster, to use the same response key for items belonging to mentally associated categories than for items belonging to nonassociated categories. The IAT has been mainly adopted to investigate attitudes (i.e., favorable or unfavorable dispositions) toward social groups, for example, related to race and ethnicity [29]. To this regard, several experimental studies have highlighted discrepancies between explicit and implicit attitudes, suggesting that the IAT can resist self-presentational forces that mask personally or socially undesirable associations, including gender and racial stereotypes and stigma, thus allowing the emergence of those associations that the individual does not wish to openly disclose [29], [31].

The IAT has found numerous applications, such as in the fields of clinical and forensic psychology, consumer research and neuromarketing [32]–[35]. For instance, consumer research applies the study of cognition and implicit preferences through the IAT to make targeted advertising and product placement more effective [36]. To date, implicit measures such as the IAT have been adopted to assess individuals' attitudes toward a specific category of AI, namely robots. Sanders *et al.* employed an IAT to study whether there is an attitude difference toward humans and robots [37]. Results showed that participants held more positive implicit associations toward other humans than robots. A similar finding is also reported in [38] and [39]. Chien *et al.* also showed that a more positive explicit attitude toward robots is achieved through direct interaction and that older adults tend to have a more negative implicit attitude than younger adults [38]. In this article, we propose to apply the IAT to investigate individuals' implicit attitudes toward a more general concept of AI. Indeed, robots are often associated with anthropomorphic characteristics [39], that may affect the implicit human perception per se. Because AI takes different forms and nowadays the interaction of humans with AI takes place mostly through nonrobot like agents (e.g., vocal assistants, chatbot), it is interesting to assume a general perspective that is free from the specific peculiarities of one or another form of AI.

III. Method

Data were collected between December 29th, 2020 and January 6th, 2021. 1139 participants were recruited through Amazon Mechanical Turk [27]. They were asked to complete a survey that consisted of two sections: 1) an IAT aimed at gathering the implicit attitudes toward AI, and 2) a self-report questionnaire to measure explicit attitudes. After cleaning up the dataset, 121 participants were excluded because they completed only the first part of the survey (IAT), while 50 participants were excluded because they did not pass one or more control-check questions intentionally placed in the questionnaire (e.g., "Please select the option B"), revealing an inaccurate response style. 139 participants were excluded because they showed a random response style to the IAT, characterized by extremely low response times. The final dataset involved 829 participants (377 females and 452 males) with an average age of 34.95 years (range 18–77 with $\sigma = 11.75$[1]) and a mean of 15.67 years of education ($\sigma = 2.09$ in the range 5–21). Geographically speaking, 457 participants resided in North America, 145 in Asia, 121 in South America, 98 in Europe, 6 in Australia, and 2 in Africa. All subjects gave their informed consent before participating in the data collection and received a compensation of 10 U.S. cents to complete the tasks. Data were collected anonymously. The present research was designed in accordance with the Declaration of Helsinki and approved by the ethics committee for psychological research at the University of Padova.

A. Study of Implicit Attitudes Through IAT

We implemented the IAT task through the Qualtrics platform and according to the original procedure proposed by Greenwald [16]. The IAT included 20 stimuli: 10 images and 10 words. Five images were related to the category "humans" and five images were related to the category "AI." Images represented humans or AI playing actions, and were tied up two by two between categories (i.e., drone for pizza delivery versus delivery boy, driverless car (DLC) versus human car driver, war drone

[1] σ indicates the standard deviation.

TABLE I
REPRESENTATION OF THE FOUR VERSIONS OF THE IAT

		Left label (key E)	Right label (key I)
Version 1	**Block 4**	humans/good	AI/bad
	Block 7	humans/bad	AI/good
Version 2	**Block 4**	AI/bad	humans/good
	Block 7	AI/good	humans/bad
Version 3	**Block 4**	humans/bad	AI/good
	Block 7	humans/good	AI/bad
Version 4	**Block 4**	AI/good	humans/bad
	Block 7	AI/bad	humans/good

versus soldier, vocal assistant versus human operator, automated production line versus human production line). Note that we chose five scenarios, where the AI is easily representable with images, as some AI products are integrated with other tools and, thus, difficult to be depicted. The "humans" images were balanced for gender (the delivery boy and the soldier were men, the human operator and the car driver were women, while in the human production line there were both men and women). As regards words, five words belonged to the category "good," so they were positive words (i.e., lovely, delight, joyful, spectacular, cheer) and five were negative words (i.e., nasty, bothersome, pain, horrible, hurtful). The IAT consisted of 7 blocks, with 20 or 40 trials each [16]. Blocks 4 and 7 were the critical ones (those we analyzed) with 40 trials each, while the others were training blocks with 20 trials each. Training blocks are designed to allow the participants to get familiar with the tasks. The structure of the IAT is reported in Fig. 1. In the first block (20 stimuli) participants were asked to classify images according to one of the two categories ("humans" vs. "AI") placed, respectively, in the upper right and upper left corner of the computer screen, pressing one of the two buttons ("E" or "I") on the keyboard corresponding to the response labels. In the second block (20 stimuli), subjects had to classify words according to one of the two categories ("good" vs. "bad"). The third block (20 stimuli) required participants to classify both images and words ("humans"/"good" and "AI"/"bad"). The same task was repeated in block four (40 stimuli), which was the one we analyzed. In the fifth block (20 stimuli), we asked participants to classify just words but, differently from the second block, the category labels were reversed in their positions ("bad" vs. "good"). The sixth block (20 stimuli) required, again, to classify both images and words, but the four categories were paired differently with respect to the third block ("humans"/"bad" and "AI"/"good"). Finally, the seventh block (40 stimuli) was the same as for block six, but it was the one we analyzed. To avoid effects due to the order of presentations of the blocks (i.e., practice effect [41]) and the positions of the labels, we created four different versions of the IAT, where the two critical blocks were inverted and the position of the labels switched. We randomly administered to each participant one of the four versions reported in Table I.

Throughout the entire task, we collected the response time (RT, or the time the participant takes to classify the stimulus) and the number of errors (number of stimuli incorrectly classified) [42]. A participant with an implicit positive attitude toward AI is expected to show faster RTs in the block where "AI" images and "good" words share the same response key. On the contrary, if the participant prefers humans to AI, having a

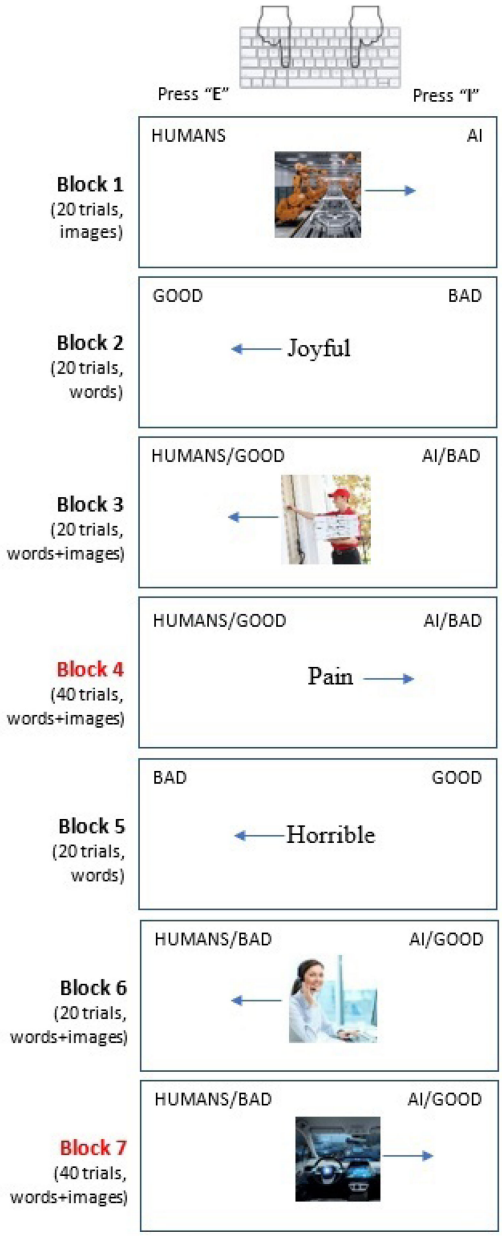

Fig. 1. Structure of the seven blocks of the IAT.

negative implicit attitude toward AI, RTs are expected to be faster when "AI" images and "bad" words share the same response key. Average response times to block 4 and 7 of the IAT were used to calculate the d-score [43]. The d-score is obtained by subtracting the average RT for the block associating "AI" and "bad" words from the average RT in the block associating "AI" and "good" words, and then dividing this difference by the inclusive standard deviation of the two blocks. The IAT's d-score was calculated via *iatgen* online [44]. A positive d-score (> 0.2) indicates a positive attitude towards humans and a negative toward AI. A negative d-score (< -0.2) indicates a preference (positive attitude) for AI over humans. A d-score between -0.2 and 0.2 indicates a neutral attitude both toward AI and humans, with no preferences for one of the two categories. Note that participants showing a 0.1 proportion of too short (< 300 ms) response times

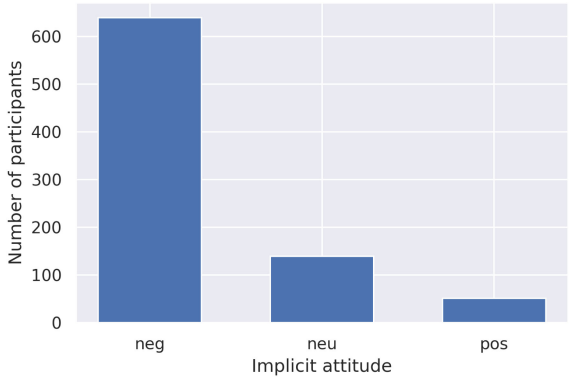

Fig. 2. Participants distribution according to their implicit attitude toward AI.

were excluded by "iatgen" [44], as it indicates the subject has responded randomly to the task.

B. Study of Explicit Attitudes Through a Self-Report Questionnaire

The second section consisted of an ad hoc questionnaire that included demographic information (e.g., age, gender, education level, occupation, and nationality) and explicit questions regarding attitudes toward AI. Explicit questions required a response on a Likert scale, ranging from 1 to 5. Questions are reported in Appendix.

IV. Data Analysis

Data analysis was conducted with Python using Pandas and Scipy Libraries [45], [46]. The significance threshold of the p-value was set to 0.05. The Pearson's product-moment correlation (r) [47] was adopted to investigate the association between age and education level and the measures of explicit and implicit attitudes; the point-biserial correlation (r_{pb}) [47] was adopted for dichotomous variables like gender. The one-way independent ANOVA [47] was run to investigate the differences in explicit and implicit attitudes between participants with an occupation outside the AI field, working in the AI field and unemployed subjects. When statistically significant results emerged, the partial eta-squared (η_p^2) was reported as an effect size. The one sample t-test (t) [47] was performed in relation to explicit measures, in order to determine whether the sample's true mean (μ) statistically differed from the central value of the Likert scale (3 = neutral attitude).

V. Results

A. Implicit Attitudes (IAT)

According to the d-score value, we classified participants in three groups: having a positive, negative, or neutral implicit attitude toward AI. The distribution of the three groups is reported in Fig. 2. The analysis highlighted that 77.10% of the participants obtained a positive d-score ($d > 0.2$), which indicates a negative implicit attitude toward AI; 6.15% obtained a negative d-score ($d < -0.2$), which means a positive attitude toward AI and 16.75% did not show any preference for humans or AI ($-0.2 < d < 0.2$).

TABLE II
Relation Between D-Score and Demographic Variables

Variable correlated with d-score	r/F	p
Age	r = 0.131	$1.63e^{-4}$
Education level	r = -0.065	0.061
Gender	r_{pb} = 0.085	0.015
'Do you work in the field of Artificial Intelligence?' (*)	$F_{(2,826)}$ = 13.829	$p = 3.93e^{-4}$ $\eta_p^2 = 0.032$

(*) the difference between three groups was tested: participants working in the field of AI ($N = 216$), participants working outside the AI field ($N = 415$) and unemployed participants ($N = 198$).

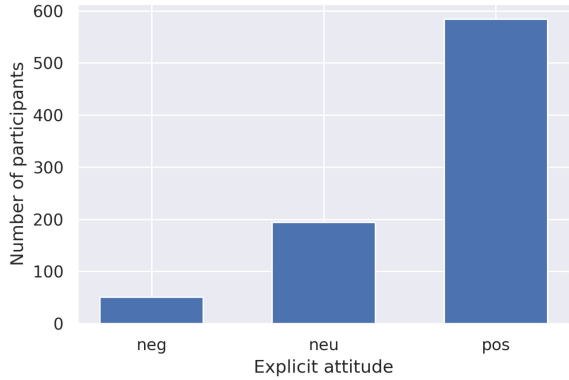

Fig. 3. Participants distribution according to their explicit attitude toward AI.

As concerns the relation between implicit attitude (d-score) and demographic variables (gender, age, education level, and occupation), it emerged that an older age was associated with a negative attitude toward AI (positive d-score). Similarly, being female was positively correlated with a negative attitude toward AI (positive d-score). An ANOVA revealed that people working in the field of AI compared to participants working out of the AI field (post hoc test: t=3.620, $p_{\text{tukey}} = 9.11e^{-4}$) and unemployed participants (post hoc test: t=5.189, $p_{\text{tukey}} = 7.98e^{-7}$) had a more positive attitude toward AI (negative d-score). Results are reported in Table II.

B. Explicit Attitudes

From the analysis of the responses to the explicit questions, it emerged that the sample mean significantly differed from the central value of the Likert scale (3 = neutral attitude). In other words, participants expressed a positive explicit opinion toward AI in all questions. Results of the one sample t-test are reported in Table III.

The most representative question of the explicit attitude was "Which is your opinion about AI?". According to this question we classified participants into three groups: having a positive, negative, or neutral explicit attitude toward AI. The distribution of the three groups is reported in Fig. 3. Just 6.25% of the participants had a negative explicit opinion about AI (responded 1 = very negative or 2 to the Likert scale); 23.40% had a neutral explicit attitude (middle value of the Likert Scale = 3); 70.45% of the participants had a positive explicit attitude (responded 4 or 5 = very positive to the Likert scale).

We tested the correlation between the explicit attitude (measured by the question "Which is your opinion about AI?")

TABLE III
RESULTS OF THE ONE SAMLE T-TEST FOR EXPLICIT QUESTIONS

Explicit Question	*t*	*p*	Cohen's d
'Which is your opinion about Artificial Intelligence?'	$t_{(828)} = 28.82$	$p < 10e^{-5}$	d = 4.298
'How favorable are you to the current use of Artificial Intelligence?'	$t_{(828)} = 24.98$	$p < 10e^{-5}$	d = 4.127
'How scared are you of the current use of Artificial Intelligence?'	$t_{(828)} = -4.927$	$p = 1.01e^{-6}$	d = 2.233
'What do you think about the impact that Artificial Intelligence currently has on the economy?'	$t_{(828)} = 20.865$	$p < 10e^{-5}$	d = 4.068
'What do you think about the impact that Artificial Intelligence currently has on society?'	$t_{(828)} = 16.21$	$p < 10e^{-5}$	d = 3.674
'What do you think about the impact that Artificial Intelligence currently has on the quality of human life?'	$t_{(828)} = 20.598$	$p < 10e^{-5}$	d = 3.82

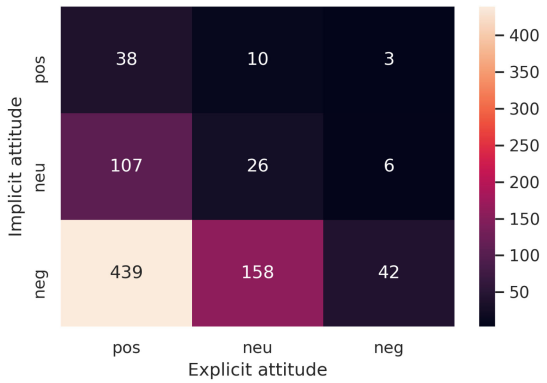

Fig. 4. Number of participants who showed a discrepancy in their attitudes from explicit to implicit.

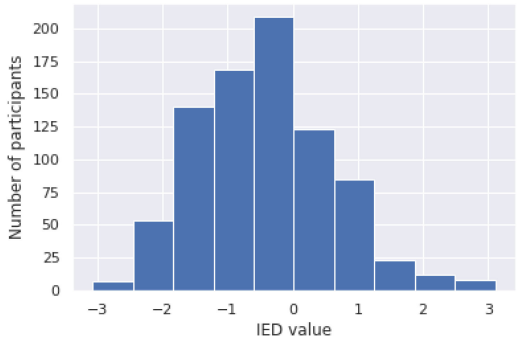

Fig. 5. Distribution of the participants according to the value of the IED.

TABLE IV
RELATION BETWEEN AI EXPLICIT OPINION AND DEMOGRAPHIC VARIABLES

Variable correlated with 'Which is your opinion about Artificial Intelligence?'	r/F	*p*
Age	r = 0.063	0.070
Education level	r = 0.084	0.016
Gender	r_{pb} = -0.070	0.043
'Do you work in the field of Artificial Intelligence?' (*)	$F_{(2,826)} = 7.829$	$p = 4.28e^{-4}$ $\eta_p^2 = 0.019$

(*) the difference between three groups was tested: participants working in the field of AI ($N = 216$), participants working outside the AI field ($N = 415$) and unemployed participants ($N = 198$).

and demographic variables (gender, age, education level, and occupation). Results are reported in Table IV. Similarly to what emerged from the analysis of the implicit attitudes, being male was correlated with a positive attitude toward AI. People working in the field of AI showed a more positive explicit attitude compared to participants working outside the AI field (post hoc test: $t = 3.916$, $p_{\text{tukey}} = 2.87e^{-4}$) and unemployed subjects (post hoc test: $t = 2.669$, $p_{\text{tukey}} = 0.021$). A higher education level was positively correlated to a positive attitude toward AI.

C. *Dissociation Between Implicit and Explicit Attitudes*

To investigate the presence of a dissociation between implicit and explicit attitudes, we created a new variable called "attitude discrepancy" that reflects the discrepancy between the attitude that emerged from the IAT (having a positive, negative, or neutral implicit attitude towards AI) and the attitude that the subject showed according to the explicit question "Which is your opinion about AI?" (having a positive, negative, or neutral explicit attitude toward AI). The analysis of the "attitude discrepancy" revealed that 87% of the participants ($n = 723$) showed a dissociation between explicit and implicit attitude, while only 106 participants maintained a coherence between the implicit and explicit measure. Fig. 4 represents the distribution of the participants according to the direction of the discrepancy of attitude: 704 participants showed a negative direction of discrepancy from explicit to implicit, while just 19 subjects presented a positive direction of discrepancy from explicit to implicit. More in detail, 439 participants with a positive explicit attitude showed a negative implicit attitude; 107 participants with a positive explicit attitude showed a neutral implicit attitude; 158 participants with a neutral explicit attitude showed a negative implicit attitude. Just three subjects with a negative explicit attitude showed a positive implicit attitude; six participants with a negative explicit attitude showed a neutral implicit attitude; 10 subjects with a neutral explicit attitude showed a positive explicit attitude.

For each participant, the IED has been calculated as the value of the difference between the standardized explicit and implicit measures of the attitude toward AI [28]. A value of zero corresponds to the absence of dissociation between explicit and implicit attitudes. The higher the value of the IED, the greater the dissociation between explicit and implicit attitudes. As concerns the direction of the IED, positive values indicate a positive implicit and negative explicit attitude; negative values indicate a positive explicit and a negative implicit attitude. Fig. 5 represents the distribution of the participants according to the value of the IED. It is worth noting that most of the participants showed a small discrepancy, especially in the direction of explicit negative and implicit positive attitude. A more consistent discrepancy is

observed for those who have a positive explicit and a negative implicit attitude.

To explore the role of the demographic variables in the dissociation between explicit and implicit attitudes, we correlated the IED with age ($r = 0.108$, $p = 0.002$), education level ($r = -0.100, p = 0.004$), and gender ($r_{pb} = 0.096, p = 0.006$). Being older, being male and having lower educational level seems to be correlated to with a grater IED discrepancy. As concerns the occupation, the results from the one-way independent ANOVA revealed that the IED of people working in the field of AI is greater than that of people working out of the AI field and unemployed ($F_{(2,826)} = 14.234,\ p = 10e^{-5}$; post hoc test working in AI vs. unemployed: $t = 4.458, p_{\text{tukey}} = 2.80e^{-5}$; post hoc test working in AI versus working out AI: : $t = -4.927, p_{\text{tukey}} \leq 10e^{-5}$).

VI. Discussion

Although AI is increasingly present in our lives, up to now scientific research has not dealt in depth with issues concerning the attitudes and trust of users toward intelligent technologies. Researchers in the field of AI recognize the necessity of building trust between humans and AI systems, improving the user experience, enhancing interpretability, and explainability of AI algorithms, and increasing the transparency of AI methods [48]. However, the knowledge about how humans think about AI is still very poor. Without understanding how the human mind perceives AI, it is difficult to build trusted and highly accepted AI systems. The few scientific studies that have investigated the perception and attitudes of people toward AI revealed a very complex scenario, characterized by mixed feelings: people recognize the positive impact of these technologies, and at the same time they show serious concerns about privacy, security, and social-economic issues. This article contributes to shed light on users' perception of AI, studying the unconscious cognitive mechanisms underlying the attitudes toward a general concept of AI. We have collected and compared both implicit and explicit attitudes toward AI, observing a clear dissociation between them. Indeed, 85% of the participants explicitly declaring to have a positive opinion about AI turned out to have an implicit negative, or at most neutral, opinion about it, suggesting that the users' attitudes toward technology suffer from unconscious and innate biases. The magnitude of this effect seems to be related to factors like age, education, gender, and familiarity with the AI field. This phenomenon, known as IED [22], is well known in psychology, especially in social psychology, where many studies highlighted the existence of implicit biases (stereotypes, prejudices) for certain stigmatized social groups, such as African Americans, homosexual people, and women [49], [50].

One possible explanation of the IED phenomenon in relation to attitudes toward AI is the sensitivity of our brain for those stimuli recognized as "different." Indeed, our brain tends to show an implicit preference for what is perceived as similar and familiar, as opposed to dissimilar or unknown [9], [51]. For example, Eyssel *et al.* [52] investigated how robot characteristics like the voice and the gender influenced the acceptance of the human–robot interaction, finding that the acceptance increased when the gender of the robot and the user were the same and when the robot's voice was similar to that of the human being. In other words, attitudes of the people toward AI seem to pass through unconscious mechanisms, which should be considered when an AI system is built and exploited to improve AI acceptance.

More recently, light has been shed also on the consequences of IED, particularly with respect to the degree of discrepancy shown by a subject. It has been argued that some people may perceive that they have both positive and negative associations toward an object, even if one of the two associations is not endorsed or is felt to be inappropriate or wrong [28]. Thus, the person is motivated to try to control the negative automatic reaction. People with a large IED process new information related to an evaluated object more carefully, probably in the attempt to reduce the discomfort deriving from ambivalence [23]. External pressures can also contribute to amplify the ambivalence and its associated feelings. For example, valuing egalitarianism when people have a large IED in relation to racial bias (high implicit and low explicit prejudice) can contribute to create ambivalence and discomfort. All these mechanisms should be considered when new ways to improve the AI acceptance are proposed or when new AI products are placed on the market. As a concrete example, sponsoring a new technological product as something that the person cannot do without, creating a social status, can generate a conflict in those people who have an implicit negative judgment toward the use of that technology and concerns toward its implications [53], [54]. The unconscious implicit bias toward AI can also lead to more general consequences on society. For instance, a negative implicit attitude toward AI may delay or discourage the introduction of technologies that can improve the quality of life, such as DLC. Even though several analyses have demonstrated that DLC may reduce traffic fatalities by up to 94% (human error) [55], it has been shown that the acceptance of DLC decreases as the level of automation increases [56]. Moreover, a negative attitude can induce an unrealistic level of expectation (e.g., DLC must guarantee almost no fatalities) that could indefinitely delay the exposure of DLC on public roadways [57]. It should be also noted that the consequences of the negative attitudes toward AI may be different according to the different forms of AI.

Another interesting result of this study, which confirms results from previous literature [1] concerns the influence of some demographic variables on AI implicit and explicit attitudes. In both the explicit and implicit measures, females show a more negative attitude toward AI than males. However, when males show an IED, this has a greater magnitude than females. A role seems to be played also by the familiarity with the AI word, as people working in the field of AI appear to be more positive towards AI, both explicitly and implicitly. Again, when they present an IED, this is greater than the discrepancy showed by people with less familiarity with AI.

As concerns the limitations of this study, one consists of having used just images representing AI products that replace people in their activities. This might have influenced the subjects' attitudes. Thus, future studies should consider this aspect and test implicit attitudes also toward AI products that enhance humans instead of replacing them. Broadly speaking, peoples' attitudes toward AI might be totally different according to the

different scenarios in which AI acts and to the different form that AI assumes (e.g., robots vs. no-robot like agents). Thus, future studies should be focused on how explicit and implicit attitudes change according to different application scenarios and AI features, including aesthetic form, functionalities, and degree of responsibility in the task. For instance, more IATs might be run to over different contexts.

A second important limitation concerns the fact that the stimuli we used for the "humans" category include highly gendered roles (e.g., a man soldier). In this way, we did not fully control for the impact of gender. To avoid the impact of gender in the evaluation, future studies should include both genders for all roles.

A third limitation refers to the comparability of the implicit and explicit measures. The measures we chose are consistent with those in previous literature. However, future research should include explicit measures that more closely paralleled the implicit ones. For instance, instead of rating the generic concept of AI, subjects could be asked to rate how good it is to use AI (versus humans) to deliver packages or drive a car, etc., allowing to consider the sum of all the ratings. A standardized and well-established questionnaire about AI acceptance could be useful.

To conclude, research in AI should consider not only explicit opinions but also implicit mechanisms when new AI technologies are designed. Implicit methods could be used to improve trust in AI and its level of acceptance. AI should not be presented to people only through science fictions films that usually describe it as dangerous and out of human control; on the contrary, more space should be given to all the positive current applications of AI, like in the medical field, allowing the user to build solid positive associations with the AI.

The debate on whether it is positive or negative and ethical to improve people's acceptance of AI is open [58]. Some authors argue that helping people overcome the prejudices they have toward AI, by increasing trust in machines, will lead to positive implications for society. A greater acceptance of AI would imply greater use of it [59] in supporting humans when making important decisions (e.g., medical, financial decisions), improving the quality of life (e.g., nutrition, physical exercise, medical screening) and work [60]–[62]. Other authors have raised their concerns, arguing that indiscriminately favoring the acceptance of AI could lead to devastating consequences for society, especially for technologies whose misuse or abuse might turn, in the long term, to be more deleterious than beneficial [63], [64]. In other words, from an evolutionary perspective an implicit negative attitude, or a bias toward AI technologies, as happens for other domains (e.g., the loss-averse phenomenon) [65] could have a protective value for humans. It follows that in the future, with the increasing spread of AI, a regulation that considers both the risks and the benefits of accepting and trusting AI will become necessary, for example by limiting the user awareness campaigns aimed at increasing their trust in intelligent machines in specific domains (e.g., preventive medicine).

APPENDIX

Self-report questionnaire investigating explicit attitudes are as follows:

1) "Which is your opinion about AI?" (from 1 = Very negative to 5 = Very positive).
2) "How favorable are you to the current use of AI?" (from 1 = Not at all favorable to 5 = Very favorable).
3) "How scared are you of the current use of AI?" (from 1 = Not at all scared to 5 = Very scared).
4) "What do you think about the impact that AI currently has on the economy?" (from 1 = Very negative to 5 = Very positive).
5) "What do you think about the impact that AI currently has on society?" (from 1 = Very negative to 5 = Very positive).
6) "What do you think about the impact that AI currently has on the quality of human life?" (from 1 = Very negative to 5 = Very positive).

REFERENCES

[1] B. Zhang and A. Dafoe, "Artificial intelligence: American attitudes and trends," *SSRN Electron. J.*, Jan. 2019, doi: 10.2139/ssrn.3312874.
[2] W. Ertel, *Introduction to Artificial Intelligence*. Cham, Switzerland: Springer International Publishing, 2017.
[3] M. X. Zhou, G. Mark, J. Li, and H. Yang, "Trusting virtual agents: The effect of personality," *ACM Trans. Interact. Intell. Syst.*, vol. 9, no. 2–3, pp. 1–36, 2019.
[4] C. Rzepka and B. Berger, "User interaction with AI-enabled systems: A systematic review of IS research," in *Proc. ICIS*, 2018, pp. 1–17.
[5] S. (Sam) Kim, J. Kim, F. Badu-Baiden, M. Giroux, and Y. Choi, "Preference for robot service or human service in hotels? Impacts of the COVID-19 pandemic," *Int. J. Hosp. Manag.*, vol. 93, Feb. 2021, Art. no. 102795.
[6] E. Fast and E. Horvitz, "Long-term trends in the public perception of artificial intelligence," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 963–969.
[7] K. Grace, J. Salvatier, A. Dafoe, B. Zhang, and O. Evans, "Viewpoint: When will AI exceed human performance? evidence from AI experts," *J. Artif. Intell. Res.*, vol. 62, pp. 729–754, Jul. 2018.
[8] Northeastern University and Gallup Inc., "Optimism and anxiety: Views on the impacts of artificial intelligence and higher education's response," 2018.
[9] D. Byrne and D. Nelson, "Attraction as a function of attitude similarity-dissimilarity: The effect of topic importance," *Psychon. Sci.*, vol. 1, no. 1–12, pp. 93–94, Jan. 1964.
[10] European Commission, "special eurobarometer 460 - march 2017 attitudes towards the impact of digitisation and automation on daily life," 2017. doi: 10.2759/835661.
[11] C. Bartneck, T. Suzuki, T. Kanda, and T. Nomura, "The influence of people's culture and prior experiences with Aibo on their attitude towards robots," *AI Soc.*, vol. 21, no. 1–2, pp. 217–230, 2007.
[12] V. C. Müller and N. Bostrom, "Future progress in artificial intelligence: A survey of expert opinion," in *Fundamental Issues of Artificial Intelligence*, Cham, Switzerland: Springer, 2016, pp. 555–572.
[13] G. Lima, N. Grgić-Hlaca, and M. Cha, "Human perceptions on moral responsibility of AI: A case study in AI-assisted bail decision-making," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, May 2021, pp. 1–17, doi: 10.1145/3411764.3445260.
[14] M. Lee, P. Ruijten, L. Frank, Y. de Kort, and W. IJsselsteijn, "People may punish, but not blame robots," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, 2021, pp. 1–11, doi: 10.1145/3411764.3445284.
[15] R. J. Rydell, A. R. McConnell, D. M. Mackie, and L. M. Strain, "Of two minds: Forming and changing valence-inconsistent implicit and explicit attitudes," *Psychol. Sci.*, vol. 17, no. 11, pp. 954–958, 2006.
[16] A. G. Greenwald and M. R. Banaji, "Implicit social cognition: Attitudes, self-esteem, and stereotypes," *Psychol. Rev.*, vol. 102, no. 1, 1995, Art. no. 4.
[17] D. Kahneman, "A perspective on judgment and choice: Mapping bounded rationality," *Amer. Psychol.*, vol. 58, no. 9,, 2003, Art. no. 697.
[18] R. H. Fazio, T. Towles-Schwen, S. Chaiken, and Y. Trope, "Dual-process theories in social psychology," *Dual-Process Theories Soc. Psychol.*, S. Chaiken and Y. Trope, Eds., The Guilford Press, 1999.

[19] F. Strack and R. Deutsch, "Reflective and impulsive determinants of social behaviour," *Pers. Soc. Psychol. Rev.* vol. 8, no. 3, pp. 220-47, 2004, doi: 10.1207/s15327957pspr0803_1.

[20] C. D. Frith and U. Frith, "Implicit and explicit processes in social cognition," *Neuron*, vol. 60, no. 3, pp. 503–510, Nov. 2008.

[21] R. H. Fazio, "Attitudes as object-evaluation associations: Determinants, consequences, and correlates of attitude accessibility," *Ohio State University Series on Attitudes and Persuasion*, vol. 4. *Attitude strength: Antecedents and consequences*, R. E. Petty and J. A. Krosnick, eds. Lawrence Erlbaum Associates, Inc., 1995, pp. 247–282.

[22] P. Briñol, R. E. Petty, and S. C. Wheeler, "Discrepancies between explicit and implicit self-concepts: Consequences for information processing," *J. Pers. Soc. Psychol.*, vol. 91, no. 1, pp. 154–170, Jul. 2006.

[23] R. J. Rydell, A. R. McConnell, and D. M. Mackie, "Consequences of discrepant explicit and implicit attitudes: Cognitive dissonance and increased information processing," *J. Exp. Soc. Psychol.*, vol. 44, no. 6, pp. 1526–1532, Nov. 2008.

[24] B. Gawronski and G. V. Bodenhausen, "Associative and propositional processes in evaluation: An integrative review of implicit and explicit attitude change," *Psychol. Bull.*, vol. 132, no. 5, pp. 692–731, Sep. 2006.

[25] R. E. Petty and P. Briñol, "A metacognitive approach to 'implicit' and 'explicit' evaluations: Comment on Gawronski and Bodenhausen (2006)," *Psychol. Bull.*, vol. 132, no. 5, pp. 740–744, Sep. 2006.

[26] R. J. Rydell and A. R. McConnell, "Understanding implicit and explicit attitude change: A systems of reasoning analysis," *J. Pers. Soc. Psychol.*, vol. 91, no. 6, pp. 995–1008, 2006.

[27] R. E. Petty, Z. L. Tormala, P. Briñol, and W. B. G. Jarvis, "Implicit ambivalence from attitude change: An exploration of the PAST model," *J. Pers. Soc. Psychol.*, vol. 90, no. 1, pp. 21–41, Jan. 2006.

[28] I. R. Johnson, R. E. Petty, P. Briñol, and Y. H. M. See, "Persuasive message scrutiny as a function of implicit-explicit discrepancies in racial attitudes," *J. Exp. Soc. Psychol.*, vol. 70, pp. 222–234, May 2017.

[29] A. G. Greenwald, D. E. McGhee, and J. L. K. Schwartz, "Measuring individual differences in implicit cognition: The implicit association test," *J. Pers. Soc. Psychol.*, vol. 74, no. 6, pp. 1464–1480, 1998.

[30] J. E. Swanson, L.. A. Rudman, and A. G. Greenwald, "Using the implicit association test to investigate attitude-behaviour consistency for stigmatised behaviour," *Cogn. Emotion*, vol. 15, no. 2, pp. 207–230, Mar. 2001.

[31] L. A. Rudman, A. G. Greenwald, and D. E. McGhee, "Implicit self-concept and evaluative implicit gender stereotypes: Self and ingroup share desirable traits," *Pers. Soc. Psychol. Bull.*, vol. 27, no. 9, pp. 1164–1178, 2001.

[32] A. Roefs *et al.*, "Implicit measures of association in psychopathology research," *Psychol. Bull.*, vol. 137, no. 1, pp. 149–193, Jan. 2011.

[33] S. Agosta and G. Sartori, "The autobiographical IAT: A review," *Front. Psychol.*, vol. 4, no. 519, 2013, doi: 10.3389/fpsyg.2013.00519.

[34] D. Maison, A. G. Greenwald, and R. H. Bruin, "Predictive validity of the implicit association test in studies of brands, consumer attitudes, and behavior," *J. Consum. Psychol.*, vol. 14, no. 4, pp. 405–415, Jan. 2004.

[35] M. Monaro, P. Negri, F. Zecchinato, L. Gamberini, and G. Sartori, "Mouse tracking IAT in customer research: An investigation of users' implicit attitudes towards social networks," in *Proc. Int. Conf. Intell. Hum. Syst. Integration*, 2021, pp. 691–696.

[36] B. Gibson, C. Redker, and I. Zimmerman, "Conscious and nonconscious effects of product placement: Brand recall and active persuasion knowledge affect brand attitudes and brand self-identification differently," *Psychol. Popular Media Culture*, vol. 3, no. 1, pp. 19–37, Jan. 2014.

[37] T. L. Sanders, K. E. Schafer, W. Volante, A. Reardon, and P. A. Hancock, "Implicit attitudes toward robots," *Proc. Hum. Factors Ergon. Soc. Annu. Meeting*, vol. 60, no. 1, pp. 1746–1749, Sep. 2016.

[38] S.-E. Chien *et al.*, "Age difference in perceived ease of use, curiosity, and implicit negative attitude toward robots," *ACM Trans. Hum.-Robot Interact.*, vol. 8, no. 2, pp. 1–19, Jun. 2019.

[39] N. Spatola and O. A. Wudarczyk, "Implicit attitudes towards robots predict explicit attitudes, semantic distance between robots and humans, anthropomorphism, and prosocial behavior: From attitudes to human-robot interaction," *Int. J. Soc. Robot.*, vol. 13, no. 5, pp. 1149–1159, Oct. 2020.

[40] Amazon, "Amazon mechanical turk," Accessed: Mar. 12, 2021. [Online]. Available: https://www.mturk.com/

[41] K. DUFF *et al.*, "Practice effects in the prediction of long-term cognitive outcome in three patient samples: A novel prognostic index," *Arch. Clin. Neuropsychol.*, vol. 22, no. 1, pp. 15–24, Jan. 2007.

[42] B. A. Nosek, A. G. Greenwald, and M. R. Banaji, "The implicit association test at age 7: A methodological and conceptual review," *Autom. Process. Soc. Think. Behav.*, vol. 4, pp. 265–292, 2007.

[43] A. G. Greenwald, B. A. Nosek, and M. R. Banaji, "Understanding and using the implicit association test: I. An improved scoring algorithm," *J. Pers. Soc. Psychol.*, vol. 85, no. 2, pp. 197–216, 2003.

[44] T. P. Carpenter *et al.*, "Survey-software implicit association tests: A methodological and empirical analysis," *Behav. Res. Methods*, vol. 51, no. 5, pp. 2194–2208, 2019.

[45] W. McKinney, "Data structures for statistical computing in python," in *Proc. 9th Python Sci. Conf.*, 2010, pp. 56–61, doi: 10.25080/Majora-92bf1922-00a.

[46] C. R. Harris *et al.*, "Array programming with NumPy," *Nature*, vol. 585, no. 7825, pp. 357–362, Sep. 2020.

[47] A. P. Field, M. Jeremy, and Z. Field, *Discovering Statistics Using R*. London, U.K.: SAGE Publications Ltd, 2012.

[48] J. Y. C. Chen, F. O. Flemisch, J. B. Lyons, and M. A. Neerincx, "Guest editorial: Agent and system transparency," *IEEE Trans. Hum.- Mach. Syst.*, vol. 50, no. 3, pp. 189–193, Jun. 2020.

[49] A.-K. Newheiser and K. R. Olson, "White and black American children's implicit intergroup bias," *J. Exp. Soc. Psychol.*, vol. 48, no. 1, pp. 264–270, Jan. 2012.

[50] E. Hehman, C. M. Carpinella, K. L. Johnson, J. B. Leitner, and J. B. Freeman, "Early processing of gendered facial cues predicts the electoral success of female politicians," *Soc. Psychol. Pers. Sci.*, vol. 5, no. 7, pp. 815–824, Sep. 2014.

[51] R. B. Zajonc, "Feeling and thinking: Preferences need no inferences," *Amer. Psychol.*, vol. 35, no. 2, pp. 151–175, 1980.

[52] F. Eyssel, D. Kuchenbrandt, S. Bobinger, L. de Ruiter, and F. Hegel, 'If you sound like me, you must be more human'," in *Proc. 7th Annu. ACM/IEEE Int. Conf. Hum.-Robot Interact.*, 2012, pp. 125–126, doi: 10.1145/2157689.2157717.

[53] S. Lahlou, "Identity, social status, privacy and face-keeping in digital society," *Soc. Sci. Inf.*, vol. 47, no. 3, pp. 299–330, Sep. 2008.

[54] Y. Z. Özcan and A. Koçak, "Research note: A need or a status symbol?," *Eur. J. Commun.*, vol. 18, no. 2, pp. 241–254, Jun. 2003.

[55] National Highway Traffic Safety Administration and U. S. Department of Transportation, "Automated Driving System: A vision for safety," Tech. Rep., 2017.

[56] C. Rödel, S. Stadler, A. Meschtscherjakov, and M. Tscheligi, "Towards autonomous cars," in *Proc. 6th Int. Conf. Automot. User Interfaces Interactive Veh. Appl.*, 2014, pp. 1–8, doi: 10.1145/2667317.2667330.

[57] M. S. Blumenthal, L. Fraade-Blanar, R. Best, and J. L. Irwin, "Safe enough: Approaches to assessing acceptable safety for automated vehicles," 2020. [Online]. Available: https://www.rand.org/pubs/research_reports/RRA569-1.html

[58] O. Gillath, T. Ai, M. S. Branicky, S. Keshmiri, R. B. Davison, and R. Spaulding, "Attachment and trust in artificial intelligence," *Comput. Hum. Behav.*, vol. 115, Feb. 2021, Art. no. 106607.

[59] D. Harrison McKnight, V. Choudhury, and C. Kacmar, "The impact of initial consumer trust on intentions to transact with a web site: A trust building model," *J. Strategic Inf. Syst.*, vol. 11, no. 3-4, pp. 297–323, Dec. 2002.

[60] J. Li, Y. Zhou, J. Yao, and X. Liu, "An empirical investigation of trust in AI in a Chinese petrochemical enterprise based on institutional theory," *Sci. Rep.*, vol. 11, no. 1, Dec. 2021, Art. no. 13564.

[61] F. Gille, A. Jobin, and M. Ienca, "What we talk about when we talk about trust: Theory of trust for AI in healthcare," *Intell. Med.*, vol. 2, Nov. 2020, Art. no. 100001.

[62] E. LaRosa and D. Danks, "Impacts on trust of healthcare AI," in *Proc. AAAI/ACM Conf. AI, Ethics, Soc.*, Dec. 2018, pp. 210–215, doi: 10.1145/3278721.3278771.

[63] A. Jacovi, A. Marasović, T. Miller, and Y. Goldberg, "Formalizing trust in artificial intelligence," in *Proc. ACM Conf. Fairness, Accountability, Transparency*, Mar. 2021, pp. 624–635, doi: 10.1145/3442188.3445923.

[64] K. H. Keskinbora, "Medical ethics considerations on artificial intelligence," *J. Clin. Neurosci.*, vol. 64, pp. 277–282, Jun. 2019.

[65] Y. J. Li, D. T. Kenrick, V. Griskevicius, and S. L. Neuberg, "Economic decision biases and fundamental motivations: How mating and self-protection alter loss aversion," *J. Pers. Soc. Psychol.*, vol. 102, no. 3, pp. 550–561, 2012.