Assessing the impact of cognitive biases in AI project development

Chloé Bernault, Sara Juan, Alexandra Delmas, Jean-Marc Andre, Marc Rodier, Ikram Chraibi Kaadoud

Assessing the impact of cognitive biases in AI project development

Chloé Bernault[1][0000−0001−7906−4768], Sara Juan[1][0000−0002−6478−1745], Alexandra Delmas[2][0009−0004−0344−0076], Jean-Marc Andre*[1][0000−0001−9844−4694], Marc Rodier[3][0000−0002−1273−0735], and Ikram Chraibi Kaadoud**[4][0000−0001−8393−1262]

[1] ENSC-Bordeaux INP, IMS, UMR CNRS 5218, Bordeaux, France
[2] Onepoint - R&D Department, Bordeaux, France
[3] IBM - University chair "Sciences et Technologies Cognitiques" in ENSC, Bordeaux, France
[4] IMT Atlantique, Lab-STICC, UMR CNRS 6285, F-29238 Brest, France
*jean-marc.andre@ensc.fr, **ikram.chraibi-kaadoud@imt-atlantique.fr

Abstract. Biases are a major issue in the field of Artificial Intelligence (AI). They can come from the data, be algorithmic or cognitive. If the first two types of biases are studied in the literature, few works focus on the last type, even though the task of designing AI systems is conducive to the emergence of cognitive biases. To address this gap, we propose a study on the impact of cognitive biases during the development cycle of AI projects. Our study focuses on six cognitive biases selected for their impact on ideation and development processes: Conformity, Confirmation, Illusory correlation, Measurement, Presentation, and Normality. Our major contribution is the realization of a cognitive bias awareness tool, in the form of a mind map, for AI professionals that address the impact of cognitive biases at each stage of an AI project. This tool was evaluated through semi-structured interviews and Technology Acceptance Model (TAM) questionnaires. User testing shows that (i) the majority admitted to being more aware of cognitive biases in their work thanks to our tool, (ii) the mind map would improve the quality of their decisions, their confidence in their realization, and their satisfaction with the work done, which impact directly their performance and efficiency, (iii) the mind map was well received by the professionals, who appropriated it by planning how to integrate it into their current work process: for awareness-raising purposes for the onboarding process of new employees and to develop reflexes in their work to question their decision-making

Keywords: cognitive bias · AI project development · awareness · user-centered design

1 Introduction

Systems using Artificial Intelligence (AI) are taking a prominent place in our lives and in businesses. These systems have become increasingly complex and

powerful and have real implications in many major fields: health [37], biodiversity [43], justice [41] and even banking [31]. Yet some AI can lead to discriminatory practices related to gender or ethnicity [19,20]. The question of biases has thus become a major issue in AI. Their identification, management, and reduction, when possible, raise several technical, human, societal, and ethical issues related to the application of deep learning algorithms [11,24,39,35]. Many research in deep learning and human-IA interaction study the impact of humans on AI systems [6,35,7] during their design, development, and management, and conversely the impact of these systems on humans [16,25,29]. Among them, two strategies stand out [27]: those focused on data and those focused on algorithms. However, few studies examine the AI development-cognitive biases link [7].
Our work, in the area of human-IA interaction, aims to fill this gap by focusing on the people involved in developing and designing AI projects and their cognitive biases. Many definitions of the concept of cognitive bias exist [47,8,33,4,35]. In the context of our work, we choose to align ourselves with the definition of [7] which presents "*A cognitive bias such as a systematic deviation of logical and rational thinking from reality*". This neutral definition (neither positive or negative) allows an objective approach of the subject. Cognitive biases are human, systematic, and universal. They are necessary for human reasoning to maintain consistency and to help fill in gaps when faced with the unknown. They also allow an individual to make a decision quickly according to his/her experience, his/her cognitive state at the time (mental load, state of fatigue, etc.), and the context. We define the development and design of AI projects as the following steps that lead to the completion of an AI project [1]: design, code implementation, testing, and production. In our work, we will refer to all of these steps as AI project development. We will call all actors in an AI project who can have an impact on the development stages of an AI system as AI professionals: AI researchers, managers, data scientists, data analysts, data architects, data engineers, and AI developers.

Our multidisciplinary work, in the field of Human-AI interaction at the intersection of cognitive science, and AI project management, is a continuation of studies conducted on the evaluation of the sensitivity of AI actors to cognitive biases [7] and the impact of the latter on intelligent systems. We question the impact of humans on AI systems during their design. Note that in our study, we make no distinction between AI systems. Our study focuses on all types of AI projects, not only those that exhibit discriminating or biased behavior towards a population.

Our research questions are associated with the following hypotheses:
- H1- the cognitive biases of AI professionals impact the projects these individuals work on
- H2- these individuals are unaware of their own cognitive biases
- H3- it is possible to create more ethical AI through raising awareness of cognitive biases among AI actors.

To test these hypotheses, our study focuses on six cognitive biases in particular, selected after a literature review for their impact on ideation and development processes [27,7]: (i) Conformity bias (ii) Confirmation bias (iii) Illusory correlation bias, (iv) Measurement bias (v) Presentation bias, and (vi) Normality bias.

Our major contribution is the realization of a mind map as a tool to raise awareness of these cognitive biases for AI professionals that addresses the impact of these biases at each stage of the AI project life cycle. This mind map was evaluated as follows: (1) A two-stage semi-structured interview session: (i) first to identify work habits without addressing cognitive biases, (ii) then to assess work habits in relation to these biases after introducing the biases targeted by the study to the interviewees; (2) Qualitative evaluation of the impact of this tool on their perception of their own biases, through user testing and observation of the professional-mind map interaction.

As the topic of bias in AI systems is much debated and studied within the AI community, we would like to clarify the contribution and position of the present work: we do not seek to establish any causal effect of the relationship between cognitive biases and algorithmic biases (a clearly distinct concept in the literature defined as *"Problems related to the gathering or processing of data that might result in prejudiced decisions on the bases of demographic features such as race, sex, and so forth"* [35]), nor do we seek to propose a tool for debiasing humans and in particular the actors involved in AI. Through our work we wish to question the cognitive biases that come into play in the human work involved in the development of an AI system (whether it is biased or not) and to highlight the concept of awareness of cognitive biases for these professionals.

We organize this article as follows: Section 2 presents related work in the area of bias analysis in AI and the methods used in prevention. Section 3 introduces the 6 cognitive biases targeted by this study. Section 4 describes the mind map, our awareness tool. In section 5, we detail the evaluations carried out: the methodology followed and the associated results. We discuss these findings and the work of this study in section 6, before concluding with prospective work in section 7. Fig. 1 presents a schematic representation of the hypothetical link between the cognitive biases of AI professionals and the possible impact on the AI systems they work on.

2 Related work

There are several types of biases impacting AI systems: cognitive biases, algorithmic biases, and biases related to the data sets [34,45]. The latter can threaten the fairness of the system for example by systematically giving advantages to privileged groups and systematically giving disadvantages to non-privileged groups [3].

Because of the multiplicity of biases and their sources, it is difficult to successfully take them all into consideration and avoid them all [10]. Nevertheless, there are tools [3] and methodologies [15,40,27,38] that can be implemented

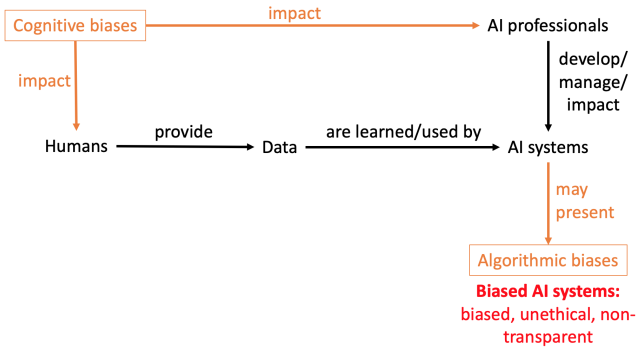


Fig. 1: Schematic representation of the hypothetical link between cognitive biases of AI professionals and the impact on AI systems.

to detect them and limit their negative consequences. For example, IBM offers "AI Fairness 360", a toolkit available in open source that ensures the fairness of an algorithm. It contains several fairness metrics for data sets and models and industry-specific tutorials to allow data scientists and others to choose the most appropriate tool for their problems. In particular, it allows to detect biases present in data sets or to evaluate the fairness of the models used [3]. As for methodologies, [15] propose a datasheet that provides a list of questions designed to obtain information about data sets. Each data set would therefore be accompanied by this datasheet to ensure transparency of the database so that users can make informed choices about how to use the data set. It would include information about its composition, collection process and recommended uses. Similarly, [27] propose to create datasheets summarizing the methods of creation, characteristics, and motivations of the data set.

There is also the SMACTR method for "Scoping, Mapping, Artifact Collection, Testing, Reflection", which allows determining the dangerous consequences that algorithms can bring before their deployment. It is an internal auditing system where developers are held accountable at each step by writing a report [40]. Finally, [38] propose a machine learning model that could handle multiple definitions of fairness and apply them. To do this, they harmonize two machine learning techniques, privileged learning [46] and distribution matching [42], to ensure that privileged characteristics such as race or gender will be information that is only used and available when the machine learning algorithms are being trained (and not in the testing or launching phase).

We have found that while the influence of human characteristics of designers on programs is known [25,48,29], there are few studies examining how cognitive biases can be concretely illustrated in the AI design process [21]. In particular, we highlight the following two works. The first work of [7], specifically addressed the issue of cognitive biases among AI professionals through a questionnaire. This study measured the sensitivity of AI professionals to three cognitive biases: conformity bias, confirmation bias, and illusory correlation bias. However, the

authors do not propose any tool or framework. In the second work of [44], the authors addressed the similar-to-me bias and stereotype bias in the context of the realization of a model of the interaction between the Human Ressources manager and the AI developer, in the context of the development of an AI system in the recruitment domain. This study, through interviews with 10 managers from New Zealand and Australia, examines how the cognitive biases of Human Ressources managers and developers lead to the development of biased AI. This model is specific to the recruitment domain and focuses on one part of the population involved in AI projects: managers and developers.

It is clear that if some works emerge in order to understand and detect the impact of cognitive biases, few approach the prism of human factors and focus on the whole life cycle of AI with different profiles of professionals (i.e. other than developers and managers). None, to our knowledge, proposes an awareness tool for the actors of AI systems design. We propose to carry out a study dedicated to this particular topic.

3 The 6 studied cognitive biases

For a given project, the cognitive biases of the project actors can intervene in every decision making [5]: from the choice of the data, to the processing that the data must undergo, to the choice of the user interface at the end of the project or even the way to represent the information [45]. Theoretically, to have a complete study about cognitive biases impacts on AI project development, we should take into account all of the 200 biases already identified [28] in the literature. However, this is impossible without conducting an overly long or complex study.

In our work, we chose to focus on six biases that have a theoretically strong impact among AI professionals and that intervene at different stages of the AI system design chain. We chose to study the **conformity bias** which consists in abandoning one's own opinion to conform to the general opinion, consciously or not [36]. We also address the **confirmation bias**, which is a tendency to look for evidence to support/confirm a diagnosis rather than to refute it [32] and finally, the **illusory correlation bias**, which consists of trying to establish/find a correlation between two variables that are nevertheless independent [17]. We chose to work on these first three biases following the work of [7] who demonstrated that they had a singular and important impact on the AIs developed. On top of that, conformity and confirmation biases come into play during the personal choices of the different actors and during interactions within the team [22]. Placing the human being at the center of our reflection and thinking particularly about human-machine interactions, it seemed to us coherent and essential to take into account the interactions within the team and not only the code produced by the individuals. This is why we chose to study the **presentation bias** which consists in influencing the perception of information by a user according to the way it is presented [27]. This bias can occur: (i) during exchanges between members of the same team who do not always have the same qualifications or the same knowledge and (ii) also during exchanges between the machine and

the user who must understand the information given by the machine without necessarily having the context of this information.

Finally, we have studied two other biases that can occur at many stages during the design of an AI. First, the **measurement bias**, which translates into subjectivity in the choice, utility and measurement of certain attributes [45,27] and which can notably lead to ethnic or gender discrimination, and second, the **normality bias**, which is a tendency to think that everything is going to happen as usual and to ignore signs indicating the opposite [2]. The role of this last bias in the field of AI is little studied. These six cognitive biases of Conformity, Confirmation, Illusory Correlation, Presentation, Measurement, and Normality are therefore at the center of our study and of the awareness tool we have created.

4 Mind Map, a cognitive bias awareness tool

In this section, we describe the design process followed to create the mind map shown in Fig. 2, as well as the mind map, its structure and functionalities, displayed in Fig. 3.

4.1 Overall design process

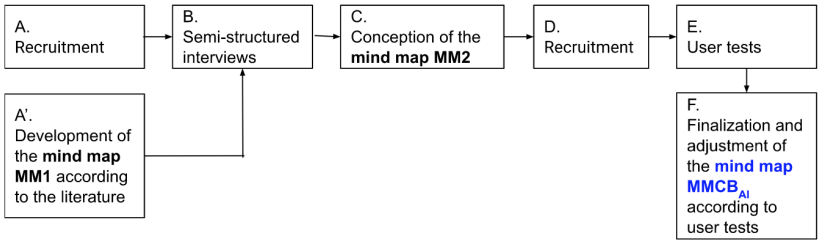


Fig. 2: Design process of the mind map

The first step of our approach was to elaborate the profile of the people we were looking for and their recruitment (Fig. 2, step A). We sought to recruit AI professionals with different profiles to participate in interviews and user testing. We identified three different professions in which to classify these actors: project managers, developers, and data scientists. It should be noted that during the recruitment process, we did not mention the subject of cognitive biases in order to avoid interviewing only professionals who were already aware of or curious about this subject. The second stage concerns the semi-structured interviews conducted (Fig. 2, stage B). We wanted to have a balance of the three professions (4 data scientists, 5 engineers, 5 project managers) so that the tool would be intended for everyone, with no disadvantaged or less well-considered professions. Thanks to these interviews, we were able to assess the sensitivity of each of the

interviewees to cognitive biases and their impact on their work. Then we also briefly presented them the first mind map that represents the stages of AI project design according to the literature and which we will refer to as MM1. This step is described in detail later in section 5.1. The third step (Fig. 2, step C) was the elaboration of the second mind map, that we will refer to as MM2, the version of the tool that we put forward in this work. We developed this tool based on the scientific literature and the feedback and results obtained during the interviews in the second stage. The fourth step consisted in the recruitment of participants to take user tests on the mind map MM2. This step constitutes the validation phase of our tool (Fig. 2, step D). For this purpose, we conducted a new recruitment campaign to recruit participants who had not previously been involved in our study. Our objective was to create a group of two types of participants to collect different opinions: those who participated in experiments 1 and 2 (who were aware of the project) and those who participated only in experiment 2 (who are neutral to the project). The fifth step (Fig. 2, step E) consisted in conducting user tests to test the MM2 mind map, validate it, and improve it. These tests are described in detail in section 5.2. Finally, the results of the user tests allowed us to make the last modifications in order to create our final tool **MindMap for cognitive biases in AI** that we will name $\boldsymbol{MMCB_{AI}}$ (Fig. 2, step E). Let us underline two points: 1) the MM1 mind map is realized in parallel to the first recruitment step. It comes from the literature. Not detailed in this work, it is an intermediary step in obtaining the MM2 mind map, the central awareness tool of this work which has been evaluated. 2) the panels of people interviewed during the second and fifth steps are different (with some exceptions discussed later).

4.2 The mind map: A decision tree for scenarios

We developed a tool for raising awareness of cognitive biases for AI professionals, in the form of a mind map (Fig. 2, step C). It is a diagram whose objective is to reflect the functioning of thought and to visually represent the associative path of thought. It is as much a tool for visualization and representation of information as for learning new concepts [9]. Mind maps are also an effective study technique that allows for better learning performance and retention of information over time, more than if the information was present in the form of written documents [12,30].

We have thus created a mind map that takes up the design steps of a project involving statistical AI[5]. We describe here the $\boldsymbol{MMCB_{AI}}$ mind map. We have

[5] Statistical AI is a subfield of AI that exploits probabilistic graphical models to provide a framework for both (i) efficient reasoning and learning and (ii) modeling of complex domains such as in machine learning, network communication, computational biology, computer vision and robotics [13].
Symbolic AI refers to AI research methods based on high-level symbolic representations of problems, logic, and search, that are accessible and readable by humans [14].
Hybrid AI combines approaches from symbolic AI and statistical AI.

identified 9 design steps according to a study of the literature [1] and the feedback from interviews conducted in the second step (Fig. 2, step B) of the design chain:

1. Analyze the problem, i.e., understand the business issues and the applications of the problem to define the objectives and the data to use
2. Collect and process the data
3. Choose the learning algorithm
4. Develop the model
5. Train and test the model
6. Visualize and analyze the results
7. Deploy/Launch the program if necessary
8. Maintain the system
9. Scale up

Our $MMCB_{AI}$ mind map is therefore presented in the form of a decision tree with 83 nodes in total. It offers 3 scenarios: a contextualization, a tutorial and the awareness part, the heart of the tool. More precisely, the mind map is composed of 12 nodes for contextualization (accessible by clicking on the "How to use it?" button) which indicates the cognitive biases we focused on, as well as our main scientific references. The tutorial has 5 nodes. As for the awareness part, it has 63 nodes. To access it, the user must click on the "Let's go" button. Fig. 3 shows an overview of the folded mind map, in French and in English, with the 3 scenarios accessible to the user.

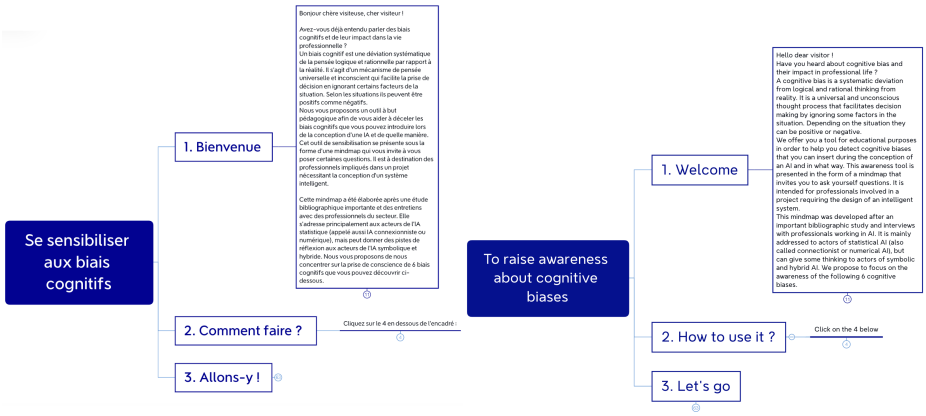


Fig. 3: Overview of the folded mind map: (a) in French, (b) in English

The awareness part focuses only on the first 6 design steps, raising questions for each of them to make users aware of the cognitive biases that can influence them. For more clarity and a better identification of the users to the problem of cognitive biases in their work, an example of a situation where cognitive biases

have negatively impacted the work of an AI actor is provided as an illustration for each step. This also allows to highlight the consequences that cognitive biases can have. Fig. 5 presents, in French and in English, an overview of the issues raised during the choice of the machine learning algorithm stage, during the implementation of the design stage and the associated example.

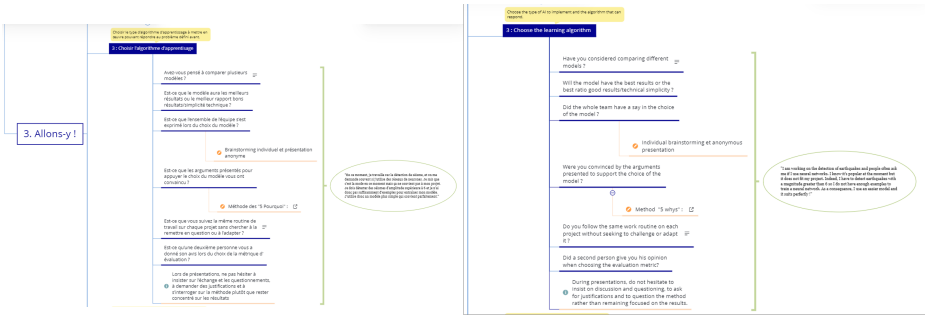


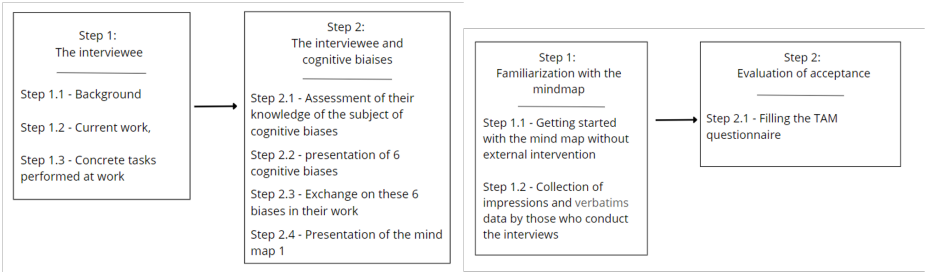
Fig. 4: Node level mind map extract: (a) "Choisir l'algorithme d'apprentissage" in French, (b) "Choose the learning algorithm"

Note that our $MMCB_{AI}$ mind map has two versions: the French version is the one used in our study. Our work was carried out in French, with a panel of French-speaking professionals, and the result is the version cited above. However, for the sake of sharing with the international community, we present an English translation for illustration purposes and to promote the reproducibility of this study. Here are the links to the:

- french version of $MMCB_{AI}$: https://www.xmind.net/m/m5nJVM
- english version of $MMCB_{AI}$: https://www.xmind.net/m/c8H8wb

5 Evaluations

In order to evaluate the impact of cognitive biases on the AI projects development, we conducted two experimental sessions: the first one before the presentation of the awareness tool in the form of face-to-face semi-directive interviews (Fig. 2, step B), and the second one after the presentation of the tool in remote mode through a screen sharing with question and answer sessions (Fig. 2, step E). We present in the following each experiment: the methodology followed, the profile of the respondents as well as the results obtained.



(a) (b)

Fig. 5: Procedures follow during experiments: (a)Experiment 1: Semi-structured interviews; (b)Experiment 2: MM2 mind map user tests

5.1 Experiment 1: Semi-structured interviews

Methodology The purpose of the first experiment was to gather information through interviews about the reality of their practices, their knowledge of cognitive biases, and their opinions about an awareness tool. We conducted semi-structured interviews with 24 questions, 18 of them are open, thus allowing for flexibility of response. This type of interview has the advantage of focusing the discussion on specific points while leaving room for rich content of information and explanatory digressions [23]. The interviews consist of 2 phases (7 steps in total) illustrated in Fig. 5a. The first phase consists of collecting information about the interviewee (Fig. 5a, step 1): his or her background, current job, and the course of the previous day. This last point allows us to know what the interviewee does as concrete tasks and allows us to know at which moments the biases can appear, without the answer being biased by the interviewee's judgment. The second phase deals with cognitive biases. First, we assessed their knowledge on the subject (Fig. 5a, step 2.1). Then we discussed the six cognitive biases we are studying (Fig. 5a, step 2.2), and we discussed how they manifest them in their work (Fig. 5a, step 2.3): when, how, and how they try to counteract them. The last phase of the interviews is the presentation of the MM1 mind map (Fig. 5a, step 2.4), a tool for raising awareness of cognitive biases in AI developed from the literature. This phase has multiple objectives: to gather users' opinions on the tool's format, content, and acceptability.

Profile of respondents We interviewed 14 professionals working on projects involving statistical AI. Among them, we count 4 data scientists, 5 engineers, and 5 project managers. Table 1 details the profile of the 14 professionals, according to their gender, age, experience in AI development, job title, and area of expertise. Our panel consists of 2 women and 12 men, all working in statistical AI except one who works in hybrid AI[6]. They have on average more than 6

[6] Please refer to the footnote 5

Table 1: Descriptive characteristics of the participants in Experiment 1

Id subject	Gender Male/Female	Age	Experience level	Job title	Field of expertise
1	F	30-40 years	10 years	Data scientist	Statistical AI
2	M	30-40 years	4 years	Data scientist	Statistical AI
3	F	18-30 years	2 years	Data scientist	Statistical AI
4	M	40-65 years	4 years	Data scientist	Statistical AI
5	M	30-40 years	3 years	AI Engineer	Statistical AI
6	M	18-30 years	1.5 years	AI Engineer	Statistical AI
7	M	18-30 years	1 year	AI Engineer	Statistical AI
8	M	18-30 years	1.5 years	AI Engineer	Statistical AI
9	M	40-65 years	1 year	AI Engineer	Statistical AI
10	M	30-40 years	8 years	AI project manager	Statistical AI
11	M	40-65 years	10 years	AI project manager	Statistical AI
12	M	40-65 years	6 years	AI project manager	Statistical AI
13	M	40-65 years	2 years	AI project manager	Statistical AI
14	M	40-65 years	34 years	AI project manager	Hybrid AI

years of experience in the field of AI with 5 people having more than 5 years of experience and 4 people having less than 2 years of experience.

Results Thanks to the interviews, we were able to understand at which moments of the design of an intelligent system cognitive biases are most likely to appear. More precisely, they intervene at the level of the different choices to be made (selection of data and construction of the data set, choice of the model to be set up, choice of evaluation metrics, choice of hypotheses, choice of inputs and outputs in Machine Learning), and during the exchanges and the design phase of the algorithms. Moreover, the interviewees mentioned that there may also be existing biases in the pre-designed data sets. Through the second step of our directional interviews (Fig. 5a, step 2.3) we identified different methods to counter the effects of the cognitive biases shared by the interviewees themselves (remember that step 2.2 in Fig. 5a consists in presenting them the 6 cognitive biases targeted by our study): exchange with various people, both internal and external to the project, diversify the teams, follow a clear and precise methodology with feedback to learn from each other's mistakes.

A major result of this first step is that we were able to note that several people declared themselves not subject to certain biases. More precisely, out of 14 people, 4 do not think they are subject to conformity bias, 4 do not think they are subject to confirmation bias, 2 do not think they are subject to illusory correlation bias, 3 do not think they are subject to measurement bias, 1 does not think they are subject to presentation bias, 8 (57%) do not think they are subject to normality bias. Among them, 2 of which because they have little experience: they argued that they had not programmed several models up to that point, and therefore they were not used to favoring one model over another. Finally, only 6 (43%) of the 14 people interviewed thought they were subject to all the cognitive biases studied. However, thinking that one is not subject to cognitive biases is the result of a cognitive bias: the illusory superiority bias [18].
Finally, we collected various opinions on the MM1 mind map. For 64% (9 peo-

ple) of the respondents, the form of a mind map seems to be judicious, 14% (2 people) are afraid that this form is too simplistic or linear in relation to their job, and the remaining 21% (3 people) have no clear opinion. From the point of view of the content, the interviewees told us that it should contain an explanation of cognitive biases, illustrative examples where cognitive biases interfered negatively, and the resulting discriminating consequences. On the other hand, they shared with us the ways in which they think they use the mind map: for example, integrating it into on-boarding processes to raise awareness among newcomers, using it to facilitate awareness workshops but also to check the awareness of a future employee, or displaying it in the office.

5.2 Experiment 2: User testing

Methodology User tests were carried out in order to evaluate the acceptance of the MM2 mind map, the quality and formulation of the questions, and finally the sensitivity of the participants to cognitive biases. We set up a test protocol taking place remotely thanks to screen sharing. We posed the following hypotheses:
H4: people who are sensitive to cognitive biases will validate and accept the tool, **vs H5:** people who are not sensitive will refuse the tool.
To collect the participants' opinions we used a questionnaire with the Likert scale (LS). This is a psychometric tool (i.e, scale) commonly used in research that employs questionnaires to measure an attitude in individuals [26]. Thus, in the first phase (Fig. 5b, step 1.1), the participants had to take in hand the MM2 mind map, and then they had to answer orally the questions of the mind map that concerned them using a LS going from 1 to 5 with 1 meaning "Totally disagree" and 5 "Totally agree" (Fig. 5b, step 1.2). This allowed assessing the participants' sensitivity to cognitive biases. In a second phase (Fig. 5b, step 2.1), participants were asked to complete a questionnaire to assess the perceived usefulness and perceived ease of use of the tool using the Technology Acceptance Model (TAM) questionnaire [49]. This TAM questionnaire is used to attempt to predict whether an individual will use or refuse to use any computer application, corporate or consumer, based on two factors: the perceived ease of use of that application and its perceived usefulness. For this questionnaire, participants were asked to rate their response on a LS ranging from 1 to 7 with 1 meaning "Strongly disagree" and 7 meaning "Strongly agree".

Profile of respondents Our awareness tool aims to be transmitted to a large number of AI professionals of all levels and professions in order to make them aware of cognitive biases and enable them to avoid them in their work. It must therefore respond to global issues and not be customized for a single panel of people. This is why we decided to take a different panel of testers than the one used for the interviews. Our panel for this second study is composed of 8 participants, whose profiles are detailed in Table 2: 2 women and 6 men. They have an average of 3 years of experience in the field of AI with 2 people who have more than 3 years of experience and 3 who have two or less. Of these, 3 people participated in the interviews (Fig. 2, step B) and 3 were trained in cognitive biases

Table 2: Descriptive characteristics of the participants in Experiment 2

Id subject	Gender Male/Female	Experience level	Participate to experiment 1	Trained in cognitive biases
1	M	3 years	yes	no
2	M	3 years	no	yes
3	F	1 year	no	no
4	M	6 years	no	yes
5	F	2 years	yes	no
6	M	3 years	no	no
7	M	1 year	no	yes
8	M	6 years	yes	no

Table 3: TAM questionnaire results: comparison of answers from participants (Likert scale of 1 to 7) who participated in both experiments and in experiment 2 only. "Answ." stands for answers.

	Answers of 3 participants that attended both experiments	Mean	Nb answ. <= 3	Nb answ. >=5	Answers of 5 participants that attend experiment 2	Mean	Nb answ. <= 3	Nb answ. >=5
Would this tool allow you to complete your work tasks more quickly?	5 2 2	3	2	1	2 6 2 6 6	4.4	2	3
Would using this tool allow you to improve your performance at work?	6 3 6	5	1	2	5 6 1 6 3	4.2	2	3
Would using this tool allow you to improve your productivity?	5 1 2	2.7	2	1	2 6 1 6 6	4.2	2	3
Would using this tool allow you to improve your efficiency?	6 3 4	4.3	1	1	4 6 1 6 6	4.6	1	3
Would using this tool make it easier to do your job?	6 3 1	3.3	2	1	1 6 4 7 6	4.8	1	4
Will you find this tool useful in your work?	6 4 6	5.3	0	2	6 4 5 7 6	5.6	0	4
How easy is it to learn how to use this tool?	5 6 6	5.6	0	3	6 7 6 7 7	6.6	0	5
Is your interaction with this tool clear and understandable?	5 7 7	6.3	0	3	7 5 7 5 6	6	0	5
Is the tool itself clear and understandable?	7 7 6	6.6	0	3	6 7 6 6 5	6	0	5

during their studies[7]. Note that no participant both conducted the interviews and studied cognitive biases. In total, we estimate that 6 participants are more aware of cognitive biases because of their training or this study. All participants answer the questions of the MM2 mind map and test the entire tool.

Results The TAM questionnaire assesses the perceived usefulness of the tool and the perceived ease of use. We had 8 answers in total presented in Table 3.

[7] They studied at the *Ecole Nationale Supérieure de Cognitique*, known also as ENSC which is an engineering school in Bordeaux, France that aims to provide an education that places humans at the heart of its designs by blending the fields of cognitive science, human-computer interaction, and AI

Table 4: TAM questionnaire results: comparison of answers from participants (Likert scale of 1 to 7) sensitive and not sensitive to cognitive biases. "Answ." stands for answers.

	Answers of 3 participants trained in cognitive biases	Mean	Nb answ. <= 3	Nb answ. >=5	Answers of 5 participants NOT trained in cognitive biases	Mean	Nb answ. <= 3	Nb answ. >=5
Would this tool allow you to complete your work tasks more quickly?	6 2 6	4.6	1	2	5 2 2 2 6	3.4	3	2
Would using this tool allow you to improve your performance at work?	6 1 6	4.3	1	2	6 3 6 5 3	4.6	2	3
Would using this tool allow you to improve your productivity?	6 1 6	4.3	1	2	5 1 2 2 6	3.2	3	2
Would using this tool allow you to improve your efficiency?	6 1 6	4.3	1	2	6 3 4 4 6	4.6	1	2
Would using this tool make it easier to do your job?	6 4 7	5.6	0	2	6 3 1 1 6	3.4	3	2
Will you find this tool useful in your work?	4 5 7	5.3	0	2	6 4 6 6 6	5.6	0	4
How easy is it to learn how to use this tool?	6 7 7	6.6	0	3	5 6 6 6 7	6	0	5
Is your interaction with this tool clear and understandable?	5 7 6	6	0	3	5 7 7 7 5	6.2	0	5
Is the tool itself clear and understandable?	5 6 5	5.3	0	3	7 7 6 6 6	6.4	0	5

•**Concerning the usefulness of the MM2 mind map in their work**: For 62.5% of the participants (5 people) the MM2 mind map would be very useful in their work (6 or 7 on the LS). For the other 3 people, the MM2 mind map would perhaps be useful (4 or 5 on the LS). Above all, it would improve the performance and efficiency of AI professionals with an average of 4.5 out of 7.
•**Concerning the contribution of the mind map to facilitate work**: For two people, the MM2 mind map would not facilitate their work at all (1 out of 7 on the LS) because it requires constant back and forth between the work done and the mind map. 50% of the participants think that the MM2 mind map will slow down their work and will not improve their productivity even if it can be very useful.
•**Regarding perceived ease of use and ease of learning to use:** Regarding perceived ease of use, 87.5% of the participants (7 people) found the tool clear and understandable (6 or 7 on the LS). Of these 87.5%, 42% even thought it was extremely clear and understandable (7 on the LS). The statistics are the same for the ease of learning to use. However, for only 62.5% of the participants (5 people), the interaction with the tool is clear and understandable.

Regarding the acceptance of our tool according to the sensitivity to cognitive biases, we noted the following:
(i) Table 3 showed that having participated in the interviews (Experiment 1) does not influence either the perceived relevance/usefulness of the tool or the

ease of picking up and using the tool. Similarly, not having participated in the interviews had no influence.
(ii) Table 4 showed that being trained in cognitive biases does not influence either the perceived relevance/usefulness of the tool or the perceived ease of use of the tool. Similarly, the fact of not having received such training has no influence.
In the current context, we were unable to confirm or refute our hypotheses H4 and H5. However, since the sample on which we rely is small, we can question this similarity in results.

Finally, thanks to the different verbatims of the participants, we were able to draw some remarks from these interviews concerning the MM2 mind map. First, the use of a mind map was not innate in all testers. Indeed, during the course of the mind map by the users, we noticed some misunderstandings on certain questions following reflections such as "I am not sure I understand" or "What does it mean?". Some of these misunderstandings could have been solved by reading the optional information accessible through a button on the side of the question. However, not all users knew that it was possible to click and get more content. Second, there were some misunderstandings related to the wording of the questions, which were sometimes too vague or too repetitive according to the interviewees because some questions were asked in two sections at the same time. This is the case, for example, of the question "Does the model work for all subgroups differentiated by socio-demographic characteristics?" which is present both in section 5, "Training and Testing the model" and in section 6, "Visualizing and Analyzing the results". We chose to keep this repetition because it ensures that the mind map user reads it, even if they skip a step. These are questions that are important to ask at different times to ensure that no bias is inadvertently introduced during the development of the tool.

The feedback from users allowed us to develop the mind map from its MM2 version to its final version $MMCB_{AI}$ presented in section 4.2. Compared to MM2, this last version contains: (i) a tutorial visible from the opening of the tool so that users can learn how to use this tool; (ii) a description of the project, as well as the main sources used to inform the user and give him confidence; (iii) clarified questions that were misunderstood by the majority of the participants by rephrasing them for greater clarity.

6 Discussion

In this paper, we address the issue of assessing the impact of cognitive biases in the development of AI projects. We focus our study on AI professionals who intervene in the life cycle of an AI project, whether they are developers, data scientists or managers, with strong technical knowledge or not. The main contribution of our work is to propose a tool to raise awareness of cognitive biases for this type of population in the form of a mind map that we have named $MMCB_{AI}$. This mind map was obtained after two experiments (Fig. 2) which allowed us to 1) study the sensitivity of the panel of participants to the notion of cognitive bias, 2) collect their opinion on a tool in the form of a mind map and

the criteria associated with the acceptance of this tool, and finally, 3) to design a final tool accepted and adapted to the needs of the AI professionals interviewed.

In more detail, Experiment 1(Fig. 2, step B) collected the testimony of interviewees on the potential impact of their cognitive biases on their job and thus the AI projects they have already conducted. In terms of results, 85% of the participants (12 people) identified at least one moment when cognitive biases impacted one of their projects or that of a colleague, which confirmed hypothesis H1. However, although the term bias was known by all participants, we also found that 43% (6 participants) considered themselves not subject to cognitive bias. On the other hand, 57% of them (8 interviewees) mentioned that one of the challenges of their work in carrying out their projects was to detect and correct the biases present in the data. While these results did not confirm or deny the H2 hypothesis, it is important to note that our participants were more likely to think that biases came either from the data or from other collaborators. The first point can be explained by the fact that AI professionals and especially technical profiles are probably more aware of technical and data-related biases in view of the development of techniques, works, and software libraries with the vocation of making ethical algorithms or unbiased data sets [10,38,3,40,15,27]. Therefore, it is easier for them to think about these biases than cognitive ones. Regarding the second point, let us also point out that, in more detail, 8 of the 14 (57%) interviewees think that they are not subject to at least one of the six cognitive biases presented. This reaction, itself the result of the illusory superiority cognitive bias, and more globally the results of this first experiment, have reinforced the need to make AI professionals aware of the topic of cognitive biases in their work.

Concerning hypothesis H3, it is not possible to deny or confirm it within the framework of our study, because we did not have enough time to measure the impact of the awareness of cognitive biases on the quality of the work of AI professionals from an ethical point of view (let us emphasize that the project was carried out in 4 months), however, the results of experiment 2 show that an approach like ours would be appreciated and useful for these professionals. Experiment 2 showed that although there are areas of improvement in the ergonomics of the tool to facilitate interaction with it, none of the 8 participants interviewed expressed an unfavorable opinion on the usefulness of the MM2 mind map (62.5%, i.e. 5 people, even declared that it would be very useful). According to them, our tool would improve the quality of their decisions, confidence in their realization, and satisfaction with the work done, which would directly impact their performance and efficiency. They even point out that the mind map could potentially be a cost-saving tool for companies since by integrating work on identifying and understanding cognitive biases and their impact during the life cycle of a project, it would allow them to anticipate and avoid more complex modifications afterward, and therefore additional costs for maintenance or correction of AI projects.

Finally, it should be noted that the principle of integrating a tool to raise awareness of cognitive biases was well received by the professionals who appro-

priated it by planning how to integrate it into their current work processes: the panels of participants in both experiments think that it would be interesting to use the mind map for awareness purposes, for example when new employees start work, and to develop reflexes in their work to question their decision-making.

Concerning the field of application of our tool, since 13 people interviewed came from the field of statistical AI, 1 person from the field of hybrid AI and none from the field of symbolic AI, we believe that this tool was designed more specifically for people working in the field of statistical AI, even if it can participate in raising awareness of actors in hybrid AI. This reflection was confirmed during a user test carried out by a person working in symbolic AI, who did not recognize himself in the questions raised by the design stages of a project involving AI.

In conclusion, more than an evaluation, our tool seeks to reinforce sensitivity by making AI professionals aware of the biases they may introduce in their work. It is important to make AI professionals aware of all types of bias, whether data-based, algorithmic, or cognitive. Indeed, a biased algorithm, regardless of the origin of these biases, has impacts on its users. The latter will potentially be led to make biased decisions and thus generate even more biased data for the training of future AI. This then creates a user-algorithm-data feedback loop that amplifies the biases [27].

7 Conclusion

The proposed methodology and the mind map, as a tool to raise awareness of cognitive biases, contribute to the field of human-AI systems interaction through a human factor and cognitive science-based approach. We consider this work as an alliance of cognitive science and AI project management.

We showed that although the AI field suffers from the issue of bias, few professionals think about the cognitive biases they carry as impacting their work and that therefore the issue of cognitive bias awareness is a current issue in the professional world working in AI. Our methodology including professionals before and after the presentation of an awareness tool allowed us to better understand the participants, their job, and their vision of this field. By giving them a voice on this subject and the possibility to act on the tool (unlike a classic training course such as a MOOC or a school course), we encouraged the creation of a context favorable to their empowerment and the questioning of their knowledge, work habits, and prior behaviors. The interactive aspect of the mind map, by encouraging the exploration of the tool, also allowed for a better appropriation of it and a projection of the participants with it. The human-tool relationship is strengthened.

As future work areas, we think it would be interesting to increase the panel (people trained in bias and untrained, for example) to re-evaluate hypotheses H1 and H2. Another line of work is to extend our study to people in symbolic AI. We believe that a new study and a tool dedicated to the sensitization of this community should therefore be carried out. Finally, we would like to explore

the impact of a more interactive tool in the form of a website or an application which we believe will be more accessible to the international community.

To conclude, we invite future research in human-IA interaction and, more globally, the AI scientific community, AI companies and educational establishments teaching AI to open up to the fields of cognitive sciences and human factors. Proposing or carrying out training at the crossroads of these multiple domains can allow for better sensitivity to cognitive biases and above all a better consideration of the human being in all its diversity. We invite the international community to test and use our $MMCB_{AI}$ mind map available online: https://www.xmind.net/m/c8H8wb. Understanding the impact of a human's cognitive sphere on its environment would allow the design of better AI tools, more adapted, and more sensitive to the different existing cognitive profiles.

Authors contributions

Sara Juan and Chloé Bernault contributed equally to this work: conception and realization of the experiments, bibliographical research, and writing of the article. Alexandra Delmas, Marc Rodiez, and Jean-Marc Andre contributed to the experiments' design and the project's supervision. Ikram Chraibi Kaadoud contributed to the bibliographic research, the design of the experiments, the writing of the article, and the supervision of the project. All authors contributed to the revision of the manuscript, read and approved the submitted version.

References

1. Barenkamp, M., Rebstadt, J., Thomas, O.: Applications of ai in classical software engineering. AI Perspectives **2**(1), 1 (2020)
2. Baron, J., Ritov, I.: Omission bias, individual differences, and normality. Organizational behavior and human decision processes **94**(2), 74–85 (2004)
3. Bellamy, R.K., Dey, K., Hind, M., Hoffman, S.C., Houde, S., Kannan, K., Lohia, P., Martino, J., Mehta, S., Mojsilovic, A., et al.: Ai fairness 360: An extensible toolkit for detecting, understanding, and mitigating unwanted algorithmic bias. arXiv preprint arXiv:1810.01943 (2018)
4. BIAS, F.O.C.: The evolution of cognitive bias. The Handbook of Evolutionary Psychology, Volume 2: Integrations **2**, 968 (2015)
5. Bonabeau, E.: Don't trust your gut. Harvard business review **81**(5), 116–23 (2003)
6. Caliskan, A., Bryson, J.J., Narayanan, A.: Semantics derived automatically from language corpora contain human-like biases. Science **356**(6334), 183–186 (2017)
7. Cazes, M., Franiatte, N., Delmas, A., André, J., Rodier, M., Kaadoud, I.C.: Evaluation of the sensitivity of cognitive biases in the design of artificial intelligence. In: Rencontres des Jeunes Chercheurs en Intelligence Artificielle (RJCIA'21) Plate-Forme Intelligence Artificielle (PFIA'21) (2021)
8. Chapman, L.J., Chapman, J.P.: Genesis of popular but erroneous psychodiagnostic observations. journal of Abnormal Psychology **72**(3), 193 (1967)
9. Cunningham, G.E.: Mindmapping: its effects on student achievement in high school biology. The University of Texas at Austin (2006)

10. Danks, D., London, A.J.: Algorithmic bias in autonomous systems. In: Ijcai. vol. 17, pp. 4691–4697 (2017)
11. Dressel, J., Farid, H.: The accuracy, fairness, and limits of predicting recidivism. Science advances **4**(1), eaao5580 (2018)
12. Farrand, P., Hussain, F., Hennessy, E.: The efficacy of themind map'study technique. Medical education **36**(5), 426–431 (2002)
13. Frontiers in Robotics and AI: Statistical relational artificial intelligence (2018), https://www.frontiersin.org/research-topics/5640/statistical-relational-artificial-intelligence, accessed on February 23th, 2023
14. Garnelo, M., Shanahan, M.: Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. Current Opinion in Behavioral Sciences **29**, 17–23 (2019)
15. Gebru, T., Morgenstern, J., Vecchione, B., Vaughan, J.W., Wallach, H., Iii, H.D., Crawford, K.: Datasheets for datasets. Communications of the ACM **64**(12), 86–92 (2021)
16. Gordon, D.F., Desjardins, M.: Evaluation and selection of biases in machine learning. Machine learning **20**, 5–22 (1995)
17. Hamilton, D.L., Rose, T.L.: Illusory correlation and the maintenance of stereotypic beliefs. Journal of personality and social psychology **39**(5), 832 (1980)
18. Hoorens, V.: Self-enhancement and superiority biases in social comparison. European review of social psychology **4**(1), 113–139 (1993)
19. Howard, A., Borenstein, J.: The ugly truth about ourselves and our robot creations: the problem of bias and social inequity. Science and engineering ethics **24**, 1521–1536 (2018)
20. Intahchomphoo, C., Gundersen, O.E.: Artificial intelligence and race: A systematic review. Legal Information Management **20**(2), 74–84 (2020)
21. Johansen, J., Pedersen, T., Johansen, C.: Studying the transfer of biases from programmers to programs. arXiv preprint arXiv:2005.08231 (2020)
22. Kahneman, D., Lovallo, D., Sibony, O.: Before you make that big decision. Harvard Business Review (2011)
23. Lallemand, C., Gronier, G.: Méthodes de design UX: 30 méthodes fondamentales pour concevoir et évaluer les systèmes interactifs. Editions Eyrolles (2015)
24. Leavy, S.: Gender bias in artificial intelligence: The need for diversity and gender theory in machine learning. In: Proceedings of the 1st international workshop on gender equality in software engineering. pp. 14–16 (2018)
25. Lepri, B., Oliver, N., Letouzé, E., Pentland, A., Vinck, P.: Fair, transparent, and accountable algorithmic decision-making processes: The premise, the proposed solutions, and the open challenges. Philosophy & Technology **31**, 611–627 (2018)
26. Likert, R.: A technique for the measurement of attitudes. Archives of psychology (1932)
27. Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., Galstyan, A.: A survey on bias and fairness in machine learning. ACM Computing Surveys (CSUR) **54**(6), 1–35 (2021)
28. Mohanani, R., Salman, I., Turhan, B., Rodríguez, P., Ralph, P.: Cognitive biases in software engineering: a systematic mapping study. IEEE Transactions on Software Engineering **46**(12), 1318–1339 (2018)
29. Nelson, G.S.: Bias in artificial intelligence. North Carolina medical journal **80**(4), 220–222 (2019)
30. Nesbit, J.C., Adesope, O.O.: Learning with concept and knowledge maps: A meta-analysis. Review of educational research **76**(3), 413–448 (2006)

31. Neves, J.M.T.d.: The impact of artificial intelligence in banking. Ph.D. thesis, Universidade Nova de Lisboa (2022)
32. Nickerson, R.S.: Confirmation bias: A ubiquitous phenomenon in many guises. Review of general psychology **2**(2), 175–220 (1998)
33. Nissenbaum, H.: How computer systems embody values. Computer **34**(3), 120–119 (2001)
34. Norori, N., Hu, Q., Aellen, F.M., Faraci, F.D., Tzovara, A.: Addressing bias in big data and ai for health care: A call for open science. Patterns **2**(10), 100347 (2021)
35. Ntoutsi, E., Fafalios, P., Gadiraju, U., Iosifidis, V., Nejdl, W., Vidal, M.E., Ruggieri, S., Turini, F., Papadopoulos, S., Krasanakis, E., et al.: Bias in data-driven artificial intelligence systems—an introductory survey. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery **10**(3), e1356 (2020)
36. Padalia, D.: Conformity bias: A fact or an experimental artifact? Psychological Studies **59**, 223–230 (2014)
37. Panch, T., Szolovits, P., Atun, R.: Artificial intelligence, machine learning and health systems. Journal of global health **8**(2) (2018)
38. Quadrianto, N., Sharmanska, V.: Recycling privileged learning and distribution matching for fairness. Advances in neural information processing systems **30** (2017)
39. Raji, I.D., Buolamwini, J.: Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial ai products. In: Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society. pp. 429–435 (2019)
40. Raji, I.D., Smart, A., White, R.N., Mitchell, M., Gebru, T., Hutchinson, B., Smith-Loud, J., Theron, D., Barnes, P.: Closing the ai accountability gap: Defining an end-to-end framework for internal algorithmic auditing. In: Proceedings of the 2020 conference on fairness, accountability, and transparency. pp. 33–44 (2020)
41. Re, R.M., Solow-Niederman, A.: Developing artificially intelligent justice. Stan. Tech. L. Rev. **22**, 242 (2019)
42. Sharmanska, V., Quadrianto, N.: Learning from the mistakes of others: Matching errors in cross-dataset learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 3967–3975 (2016)
43. Silvestro, D., Goria, S., Sterner, T., Antonelli, A.: Improving biodiversity protection through artificial intelligence. Nature sustainability **5**(5), 415–424 (2022)
44. Soleimani, M., Intezari, A., Taskin, N., Pauleen, D.: Cognitive biases in developing biased artificial intelligence recruitment system. In: Proceedings of the 54th Hawaii International Conference on System Sciences. pp. 5091–5099 (2021)
45. Suresh, H., Guttag, J.: A framework for understanding sources of harm throughout the machine learning life cycle. In: Equity and access in algorithms, mechanisms, and optimization, pp. 1–9. Association for Computing Machinery, New York, NY, USA (2021)
46. Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. Neural networks **22**(5-6), 544–557 (2009)
47. Wason, P.C.: On the failure to eliminate hypotheses in a conceptual task. Quarterly journal of experimental psychology **12**(3), 129–140 (1960)
48. West, S.M., Whittaker, M., Crawford, K.: Discriminating systems. AI Now (2019)
49. Yves Martin, N.P.: Acceptabilité, acceptation et expérience utilisateur: évaluation et modélisation des facteurs d'adoption des produits technologiques. Ph.D. thesis, Université Rennes 2 (2018)