


# Human-Centric AI to Mitigate AI Biases: The Advent of Augmented Intelligence

Antoine Harfouche, University Paris Nanterre, France\*

 <https://orcid.org/0000-0002-0407-9217>

Bernard Quinio, University Paris Nanterre, France

Francesca Bugiotti, Paris-Saclay, CNRS, LISN, CentraleSupélec, France

## ABSTRACT

The global health crisis represents an unprecedented opportunity for the development of artificial intelligence (AI) solutions. This article aims to tackle part of the biases in artificial intelligence by implementing a human-centric AI to help decision-makers in organizations. It relies on the results of two design science research (DSR) projects: SCHOPPER and VRAILEXIA. These two design projects operationalize the human-centric AI approach with two complementary stages: 1) the first installs a human-in-loop informed design process, and 2) the second implements a usage architecture that aggregates AI and humans. The proposed framework offers many advantages such as permitting to integrate of human knowledge into the design and training of the AI, providing humans with an understandable explanation of their predictions, and driving the advent of augmented intelligence that can turn algorithms into a powerful counterweight to human decision-making errors and humans as a counterweight to AI biases.

## KEYWORDS

Algorithms Biases, Human-Centric AI, Human-In-The-Loop AI, Informed AI, Intelligence Augmentation

## INTRODUCTION

Due to black swan events in the context of global health crises (Chen et al., 2021), organizations are increasingly integrating artificial intelligence (AI) into their operations (Dwivedi et al., 2021). During a crisis the most critical goal of most organizational decisions is to effectively utilize scarce resources and improve performance (Johnson et al., 2022). With its ability to process and analyze a large volume of data, quicker than a human brain can, AI helps determine possible consequences of actions and streamlines the decision-making process (Harfouche et al., 2022).

Many AI projects have been considered failures. For example, in 2016, the chatbot Tay was introduced by Microsoft with the promise of an “AI with zero chill,” but it quickly began to make racist and derogatory remarks in response to aggressive Twitter users. On March 18th, 2018, Elaine Herzberg paid with her life due to an AI failure (Smith, 2018). She was fatally struck by an automated

DOI: 10.4018/JGIM.331755

\*Corresponding Author

This article published as an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>) which permits unrestricted use, distribution, and production in any medium, provided the author of the original work and original publication source are properly credited.

Uber test vehicle while pushing a bicycle across a four-lane road in Arizona. Many researchers have cautioned that some of these AI failures are related to the development of biased algorithms (see, e.g., Akter et al., 2022; Johnson et al., 2022; Martin, 2018; Mittelstadt et al., 2016; Ziewitz, 2015). Bias in AI can occur during data collection, AI design, training of the algorithm, and interpretation of outputs, as well as after deployment and use (Harfouche et al., 2023).

Artificial intelligence is mainly used in situations that require capturing vast amounts of data according to Akter et al. (2022), and it exhibits characteristics of human intelligence (Huang & Rust, 2021) through learning from external data (Kaplan & Haenlein, 2019). Collins et al. (2021) consider that there is an urgent need to define AI to help policymakers better identify potential threats and opportunities and orient research toward the needed frameworks. They call for an increase in the number of rigorous AI academic studies, a better and more detailed definition of AI in information systems (IS) studies, and an installation of a general process of cumulative knowledge. In this paper, we adopt the definition from Rai et al. (2019) that considers AI as “the ability of a machine to perform cognitive functions that can be associated with human minds, such as perceiving, reasoning, learning, interacting with the environment, problem-solving, decision-making, and even demonstrating creativity” (p. iii). We will consider AI as several machine learning (ML) algorithms that build a model of rules or links learned from training data (Harfouche et al., 2019). Artificial intelligence can learn from data by automatically identifying hidden patterns and building decision-making models. Most data, however, are biased (Akter et al., 2022). Naturally, ML also reflects the bias inherent in the data itself. Machine learning models can replicate and sometimes exacerbate existing biases (Harfouche et al., 2023).

We examined the following research question: How can human-centric AI mitigate AI biases and contribute to the advent of augmented intelligence?

If the challenges of past decades were associated with social phenomena of knowledge transfer and knowledge creation, the main challenge today is related to human-computer interaction, and more specifically, how to combine the abilities and knowledge of human beings with various AI algorithms. A key sustainability challenge in artificial intelligence is the need for more collaborative, transdisciplinary, and robust scientific involvement in the design of AI architecture, training of AI agents, explanations about hypothesis validation, and continuous usage of AI.

We tackled various biases by designing human-centric applications based on a collaboration between humans and AI. Our architecture designs focus on intelligence augmentation which can be defined as computers enhancing human intelligence (Jain et al., 2018). Human-centered AI can be designed to continuously collaborate and learn from human input while providing explainable and interpretable predictions (Chen et al., 2021; Johnson et al., 2022; Johnson et al., 2021; Tutan et al., 2022). Explainability is a condition for a human-centric approach. According to Horvatić and Lipić (2021), human-centric AI allows humans to control and continuously improve AI applications’ performance, robustness, fairness, accountability, transparency, and explainability.

In this paper, we fill two gaps in current research—one academic and the other managerial. On the academic side, research shows the superiority of approaches mixing human knowledge with the use of AI tools, but there is little or no approach to operationalize this mix. On the managerial side, the use of AI in companies often stagnates at the level of proof of concept (POC) or proof of value (POV) (Potelle & Leblond, 2018) due to the lack of a method to integrate tools into the operational processes. To close these two gaps, we developed a new AI approach through the design science research (DSR) methodology (Hevner et al., 2004). It relies on the results of two design science research projects: 1) Schopper, conducted between 2017 and 2020; 2) and Vrailexia, designed in 2019 and initiated in 2020.

The rest of the paper is organized as follows. The next section resumes address of four approaches to collaboration between humans and AI as described by Shrestha et al. (2019). A discussion on distinct AI biases related to the collection, data preprocessing and storing, AI design, and AI implementation and use is presented, followed by an introduction to a human-centric AI and discussion about how it

is related to augmented intelligence and how it can mitigate biases. Then resumes the methodology of the research and the related two action research projects: Schopper and Vrailexia. The research results are then presented before discussing the theoretical and managerial implications of the proposed approaches and providing a summary of conclusions.

## AI USE IN COLLABORATION WITH HUMANS

There is an abundance of research on the operational collaboration between human and AI tools (Baird & Maruping, 2021; Rai et al., 2019). While several models and approaches have been proposed, the Shrestha et al. (2019) model is of a particular interest because it was designed in the specific context of decision-making.

They compared four kinds of collaboration between humans and AI in a decision-making process (see Table 1) which are 1) full human-to-AI delegation; 2) hybrid-decision: AI-to-human; 3) hybrid-decision: human-to-AI; and 4) aggregated human–AI decision-making.

### Full Human-to-AI Delegation

In AI implementations involving full human-to-AI delegation of decision-making, according to Shrestha et al. (2019), AI makes decisions without human intervention but under full human responsibility. Using AI in this way is useful in scenarios where the interpretability of the decision-making process is less important than the prediction's replicability and speed. AI-based online fraud detection, AI-based traffic planning, real-time product recommender systems, and AI-based dynamic pricing are examples of such applications.

### Hybrid Decision-Making: AI to Human

With regard to implementations involving hybrid decision-making, humans and AI sequentially make decisions to benefit from the strengths of both while amplifying each other's weaknesses (Shrestha et al., 2019).

In the AI-to-human collaboration, again according to Shrestha et al. (2019), AI evaluates the initial set of alternatives. It rejects redundant or inappropriate ones and passes on a subset of those suitable for a human decision-maker to select from, which allows them to effectively handle situations involving a large set of alternatives. This design finds its applications mainly in crowdsourcing contests, healthcare monitoring, hiring, and loan application assessment.

**Table 1. Types of human collaboration with AI**

Type of Collaboration	Field Width	Interpretability	Speed	Replicability	Data
Full Human-to-AI Delegation	Narrow due to AI restriction	Low	Fast	High	High volume
Hybrid Decision Making: AI to Human	Narrow	High because humans make the final choice	Slow because of sequential human intervention	Low because of human intervention	High volume
Hybrid Decision Making: Human to AI	Medium	Low because AI makes the final choice	Slow because of sequential human intervention	Low because of human intervention	Medium volume
Aggregated Human–AI Decisions	Large	Medium	Slow because of bottleneck humans	Medium	May be low

Note: This table has been adapted from Shrestha et al. (2019).

## Hybrid Decision-Making: Human to AI

In the second possible hybrid structure, the human decisions can be designed as inputs to algorithmic decision-making. In this case, human decision-makers select a small set of alternatives and then pass it to the AI for evaluation and selection of the best one. This design is effective when humans have high confidence in a small set of preferred alternatives, but the decision process requires an evaluation of a large amount of data. The final decision made by AI makes the process quicker. This design applies mainly to sports analytics and in health care monitoring of bodily functions (e.g., heart rate, temperature, blood pressure).

## Aggregated Human and AI Decisions

When implementing an aggregated human and AI decision process, different elements are allocated to humans and AI based on their respective strengths and weaknesses. Individual decisions are then combined into a collective decision using an aggregation rule such as majority voting or another rule of votes. This design reduces the interdependence between humans and AI allowing for a reduction in the risk of amplification of human errors (Shrestha et al., 2019).

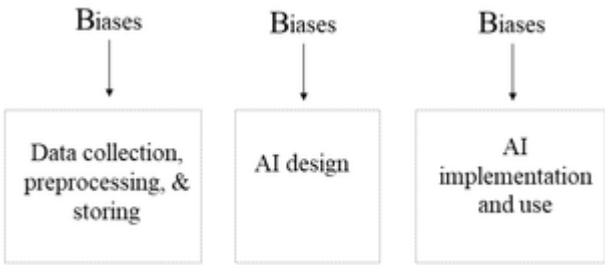
Comparing the fourth category with the three previous ones shows that the aggregation of human and AI decisions offers medium interoperability, speed, and replicability while increasing the field width. The aggregated human and AI decisions would be of high interest if the interpretability and replicability and speed of their decisions is improved.

## AI BIASES

The benefits and implications of AI are tremendous. AI is a powerful tool that can be applied to many complex problems which have not yet been successfully addressed. But AI solutions are not neutral (Akter et al., 2022; Tsamados et al., 2021). There are hidden and unchecked biases in their design (Martin, 2018; Mittelstadt et al., 2016; Ziewitz, 2015). As shown in Figure 1, biases are incorporated into the full process of AI design: 1) data collection, preprocessing, and storing 2) AI design, and 3) AI implementation and use.

When the data collected is biased, AI recommendations and decisions can yield unintended and negative consequences (Crawford & Calo, 2016; Mittelstadt et al., 2016). Training data can cause bias in two ways. First, if the training dataset does not represent a random sample from the target population, it will produce either a sample inadequacy or a sample selection bias (Akter et al., 2022). Second, if the sample was selected from an incorrect target population, it can produce an out-group homogeneity bias (Akter et al., 2022). The bias can also be related to the data preprocessing and storing. Sometimes the bias is deeply embedded in real life and consequently in the data collected. If not corrected during the data preprocessing, these datasets can exacerbate issues of inconclusive

Figure 1. Biases and AI design process



data and biased predictions. A thorough evaluation of the available datasets and their processing to mitigate biases is a key step in AI design.

The AI design and the selection of the algorithm can also be bias embedded. For example, most ML algorithms identify correlations between variables in the underlying data but without being able to identify causal relations. As such, two biases are most likely to appear: the correlation fallacy bias that confuses correlation with causation (Akter et al., 2022) and the apophenia bias that sees patterns where none exist (Mittelstadt et al., 2016). These biases are different from human cognitive ones. Indeed, according to Haselton et al. (2005), cognitive biases are mainly rooted in errors in thought processing arising from problems with attention, memory, attribution, and other human cognitive mistakes. The correlation fallacy bias and the apophenia bias are amplified when massive quantities of training data are used by mistakenly offering connections that radiate in all directions (Boyd & Crawford, 2012) which leads to inconclusive evidence.

Once AI is implemented, the associated self-learning algorithms are characterized by their degree of transparency and accountability (Buhmann & Fieseler, 2021). If certain algorithms are directly interpretable (e.g., the random forest algorithm), they need to be explained by tools such as SHAP (SHapley Additive exPlanations) (Lundberg & Lee, 2017) or LIME (Local Interpretable Model-Agnostic Explanations).

Even though AI solutions render data sets valuable, the risk of ending up with biased AI is remarkably high. Hence, there is a consensus on the need to develop human-centric AI (Nahavandi, 2019). Human-centered AI mixes technological feasibility with human perspective and knowledge in order to integrate human debiasing capabilities into the design. It supposes that in designing AI there is a need to converge algorithm capabilities with human knowledge to reduce biases while increasing interpretability and replicability and also include what is desirable for humans.

## **A HUMAN-CENTRIC AI FOR AN AUGMENTED INTELLIGENCE**

While business research is adopting the AI revolution at full speed, the design of human-oriented applications for management decision-making is not without challenges. Indeed, the design of AI applications is a multidisciplinary effort, involving extensive collaborations among data scientists, information systems specialists, and experts and users from the specific domain for which the AI is designed (Harfouche et al. 2022). In order to bring AI to its most useful state, multidisciplinary scientists must come together to establish common standards and adapt design platforms and applications. Thus, in management, integrating knowledge coming from both the field of ML and AI design with knowledge about the managerial context is imperative to mitigate AI biases (Harfouche et al., 2023).

User participation in the design of information systems is not new (Hirschheim, 1985), but in its classical approach, it was reduced to the analysis of needs and tests (Issa & Isaias, 2022). In the human-centric AI approach, participation is requested at the expert and user levels all along the process (Johnson et al. 2022; Rožanec et al. 2022). Experts know the domain well, so they must intervene in all phases of the design and the implementation processes (Johnson et al. 2022). That makes this approach similar to agile methods, but those engage the experts of the domain and not the users.

Companies' performance can be improved when AI and humans collaborate (Marnewick & Marnewick, 2020; Wilson & Daugherty, 2018). Figure 2 shows the two main interactions produced among the three main components of the AI design process: 1) Human alimentering and aggregating the AI with their knowledge and 2) AI amplifying and augmenting human knowledge. Following this motivation, the recently emerging trend in AI design is based on the exploration of human-in-the-loop (Grønsund & Aanestad, 2020; Luo et al., 2022) approaches and the development of contextual explanatory AI applications. The human-in-the-loop informed AI approach involves human knowledge

in all steps of the AI design, from data collection and data cleaning to algorithm selection, training, testing, and AI implementation and interpretation. It is similar to informed AI (Johnson et al., 2022; Karpatne et al., 2017; Von Rueden et al., 2023) but adds the deployment and the ongoing use of the informed approach. While data-driven AI purely uses data to detect patterns, informed AI considers additional knowledge and builds a second source of information (Johnson et al., 2022; Von Rueden et al., 2023).

As shown in Figure 2, humans can enhance AI models by incorporating levels of abstractions into the analysis process, while AI can augment human decision-making capabilities by analyzing possible consequences of each action and explaining the reasons behind such possible consequences. According to Wilson and Daugherty (2018), the human role starts with the design and training of the AI, and it then continues during the explanation and implementation. It proceeds in a loop to sustain the trustworthy implementation and use of the AI.

Based on the Harfouche et al. (2017) framework, the human-centric approach is composed of six stages as shown in Figure 3. The process starts with human knowledge extraction and finishes with deployment and continuous usage.

### Extract Corpus of Knowledge

In the first stage, tacit knowledge (Karpatne et al., 2017) from human experts is collected and integrated with explicit (i.e., scientific) knowledge as well as with raw data (Curtarolo et al., 2013; Faghmous et al., 2014) to help design and implement informed AI models effectively. Based on experts' knowledge from different domains and ontologies, potential features that can explain the target variable are identified and datasets are explored. Additionally, data ingestion merges data from different sources

Figure 2. The human-centric AI approach

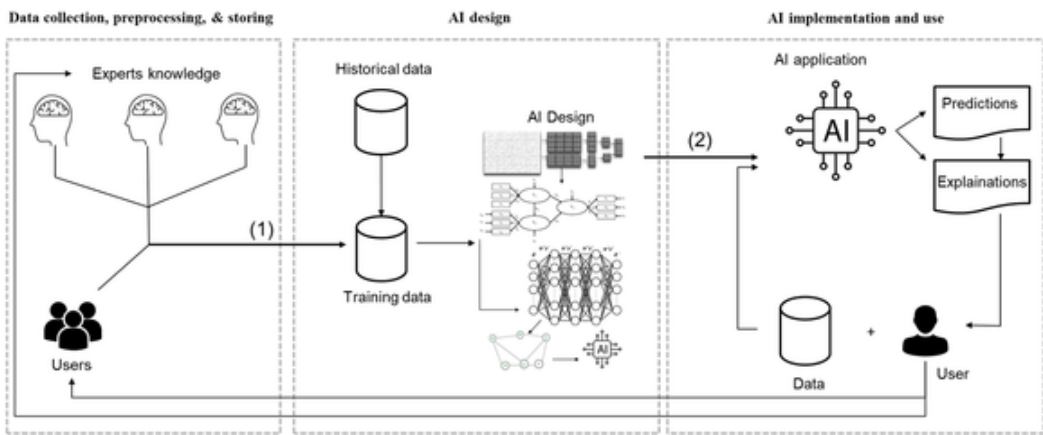
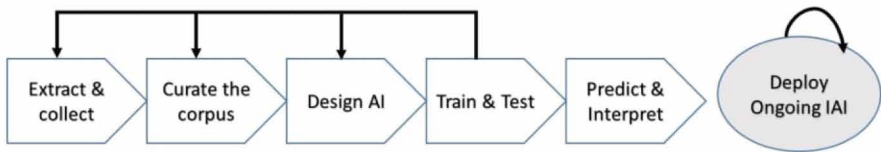


Figure 3. Six stages of human-centric AI



Note: This figure has been adapted from Harfouche et al. (2017)

and creates indices and metadata (Karpatne et al., 2017). With the help of data scientists, the process is focused also on discovering unjust biases that exist in the training data (Martin, 2018). Data sampling is employed to avoid the side effects of data imbalance (Batista & Monard, 2003), while data balancing is used by resampling and transforming the training set (Pecorelli et al., 2020).

### **Curate Content**

This stage refers to simplifying, preprocessing, and transforming raw data using expert knowledge (Harfouche et al., 2017). Through data cleansing, outliers, irrelevant variables, and observations that do not contribute to the overall model performance are removed (Diligenti et al., 2017; Yu et al., 2018). Data cleansing reduces the dimensionalities of data and helps AI algorithms converge faster (Hinton & Salakhutdinov, 2006). Training instances are then assigned a salient attribute and labeled to create the target variable by extensively using tacit human knowledge (Johnson et al., 2022).

### **Design the AI Architecture**

Determining the right AI architecture requires domain experts to deliberate on the business problem. First, feature engineering techniques are applied to extract new variables that help enhance the performance of AI models (Domingos, 2012). Then, using feature selection algorithms, a small subset of features is automatically chosen to improve model performance and reduce training time (Harfouche et al., 2017). Because employing an ensemble of different AI algorithms can generate parsimonious models (Rokach, 2010), domain experts identify multiple AI algorithms based on the problem type. Afterward, a hyperparameter search space for each algorithm is determined by domain experts. Hyperparameter optimization that uses the grid search technique is then performed to determine the best hyperparameters for each model. The final AI model maps the original features into accurate predictions of the target variable (Camacho et al., 2018).

Data scientists provide the expertise necessary to incorporate ML and design AI architectures that can solve specific scientific problematics. Their implication goes beyond simply selecting the right algorithm adapted to the specific research question. Human knowledge must be incorporated into the design of the AI to ensure that the historical reference points are appropriate and fair.

### **Train and Test**

This stage includes debugging and adjusting AI models to address overfitting and underfitting issues and selecting appropriate performance metrics (Yu et al., 2018). Overfitting happens when the trained AI model succeeds in predicting the training data with high accuracy while failing to make predictions on the test data. Underfitting occurs when the AI model does not accurately predict the training data. Overfitting is addressed by increasing the size of the training data or decreasing the model complexity. Underfitting is resolved by making the AI model more complex. In this stage, an appropriate performance metric among different key measures (e.g., accuracy, recall, AUC) is selected by human experts for accurate and reliable interpretation of results. Finally, AI models are debugged to improve and refine the results.

### **Predict and Interpret**

During the predict and interpret stage, the best AI model generated in the previous step is used to make predictions for new datasets (Camacho et al., 2018). The AI model's properties are summarized, cause-and-effect relationships are explained, and managerial insights are extracted. Employing domain knowledge and expertise in this stage is vital to explain and interpret the models accurately, validate the results, and test specific relationships among variables (Hevner et al., 2004).

### **Deploy an Ongoing Augmented Intelligence**

The collaboration between AI and humans is perpetuated sustainably in order to amplify new knowledge during this stage. The human-AI aggregation facilitates in the discovery of new ways to

make AI more efficient and less biased (Tsamados et al., 2021). The ongoing deployment of augmented intelligence can turn algorithms into powerful counterweights to human decision-making errors and human counterweights to AI biases.

## METHODOLOGY

### Design Science Research (DSR)

To achieve our research objectives, we have adopted a design science research (DSR) methodology. The reason behind this choice is due to the fact that existing IS research has been increasingly using it as a framework for designing and implementing AI-based artifacts (Adam et al., 2021). The advantage of such methodology is that it offers the ability to develop an AI artifact that can solve business and societal challenges while also contributing to knowledge creation (Hevner et al., 2004). This is in line with our research objective, which is to design new AI applications to address managerial challenges while reducing the biases that can be produced during such processes. Therefore, based on the DSR paradigm, we present in this paper two innovative AI artifacts. The first one was developed for Schopper and the second one was proposed for Vrailexia. Table 2 details the seven requirements of DSR applied to both projects.

The two projects are presented in the following subsections. Even though the focus of this article is on the design of two AI artifacts, the global context is essential to understand the similarities and specificities of these two projects.

### Project Schopper

The Schopper project aimed to produce a simulator that would be able to construct and evaluate the behavior of prehistoric man in a reconstructed immersive environment (Appendix A). It was

**Table 2. The seven requirements of DSR research**

Requirement	Details
Design as an Artifact	Two different designs were conceived based on a common framework (the six stages of human-centric AI that was adapted from Harfouche et al. [2017]).
Problem Relevance	Schopper and Vrailexia aim to tackle real complex social problems. These problems cannot be tackled with simple designs. Both projects required to rely not only on data but also on the domain experts' knowledge.
Design Evaluation	The design process included feedback steps based on assessing the AI application by potential users. Frequent meetings were organized to evaluate potential results. Improvements were proposed.
Research Contribution	The human-centric approach offered innovation possibilities, especially in tackling AI biases. New methods were introduced to enrich missing and small quantities of data. Other methods were proposed to tackle multidisciplinary collaboration issues.
Research Rigor	Two authors of this paper were implicated in the design of both projects. It was possible for the mission to keep a distance and to adopt a wider neutral view related to the implementation steps. Many seminars and meetings were organized to discuss and evaluate the reliability of the process and the results obtained.
Design and Research Process	Both projects were implemented in a process that includes six stages. Many tests were realized, and a transfer of knowledge occurred between the different tests.
Communication of Research	Both projects were designed in the context of government funding. The funding applications were positively evaluated by reviewers and both projects were awarded. Each project offers valuable technical and managerial implications for the IS community.



originally designed for a specific archeological site: The Caune de l'Arago. But its use extends to other locations. The project initiator, the CERP (Centre Européen de Recherche Préhistorique) is a prehistoric archeology research center. The CERP collaborates with two firms (one specialized in virtual reality and another one in artificial intelligence) and a research center specialized in management and Information Systems (CEROS). Two of the authors of this paper are members of the CEROS research center.

The Caune de l'Arago is in South of France, located between the Pyrenees Mountains and the Mediterranean Sea. It is a major cave site from the Lower Paleolithic. Since 1967, archeological objects (e.g., animal bones, lithic remains or industries, stone, etc.) extracted from the site have been numbered, drawn to scale, and entered in an excavation book with an identity and its spatial coordinates. They are then recorded at a laboratory in an SQL database called "Paleontological and Prehistoric Material" that was developed in the 1980s. This database contains more than 500,000 archeological objects. Archeologists have maintained this database since its creation with technologies that have evolved over time. These queries were mainly related to the extraction of quantitative data and the study of spatial repartition.

Schopper aimed to develop three different artifacts: 1) a representation of the database in virtual reality (VR) where researchers can interact with objects using a VR headset (Quinio et al., 2020); 2) an AI simulator for predicting factors based on the analysis of the database at a specific time (Grégoire et al., 2021); and 3) a representation of the valley outside the cave in VR at the desired time in connection with the AI simulator (Harfouche et al., 2023).

Twelve archaeologists engaged in this project, one of them playing a key role assigned as a full-time researcher. The potential final users of the application were researchers and students of archeology. The data scientist who implemented the design was an employee of an AI firm, a partner of the project. Two authors of this paper conducted the design and follow-up of the project.

## Project Vrailexia

Vrailexia is a project that aims to develop an AI and VR application that can increase the inclusion of dyslexic students and improve their chances of success during their academic career and integration into the labor market (Appendix B). Dyslexia is a learning disorder of written language that is linked to other learning disorders that are grouped under the umbrella of DYS. Data on dyslexic students in higher education are minimal. It is known that in secondary education, the percentage of students suffering from dyslexia is between 3% and 5%, while some studies announce nearly 8% of the global population is affected. Dyslexic students face many difficulties during their university careers resulting in a higher drop-out rates than other students. Today, most of the support for dyslexics is mainly concentrated around primary and secondary school, while few comprehensive approaches are offered for higher education.

The three-year project began in December of 2020. It is composed of 10 partners (European higher education institutions) from six different countries. Vrailexia aims to create three categories of tools: 1) a VR application that aims to test and assess students' difficulties and propose roadmaps and scenarios that allow teachers to better understand the problems of dyslexic students while immersing themselves in reconstructed environments and situations; 2) a support platform for students with dyslexia based on adaptive learning from AI— digital solution based on AI designed to propose the most appropriate strategy and tools for dyslexic students based on their own difficulties; and 3) a platform to share resources between all partners to support dyslexic students in their studies and job searches and to raise awareness of nonvisible disorders among all the actors of the educational system.

The training data was based on a survey administered in three countries (France, Italy, and Spain) and a psychometric test and other data collected through the VR application. The survey consisted of four parts: 1) a categorization of the dyslexic respondents (i.e., age, family, studies), 2) the difficulties they encounter selected from a list, 3) the tools they use to overcome their difficulty, and 4) the strategies they implement. The spell checker is an example of a tool, while

a particular notetaking method and the use of diagrams are examples of strategies adopted by dyslexic students.

The training data was complemented and improved by extracting the reasoning taxonomy of experts in dyslexia collected by qualitative interviews. These experts include psychologists, researchers, speech therapists, and heads of dyslexia-related associations.

The data scientists implicated in the project are from Italian and French universities. The final users are dyslexic students from different countries, some who were involved with beta testing.

### Distance Between the Researchers and the Object of Study

The Schopper project has been completed. Two authors of this paper were implicated in this project from its submission to its completion. They participated in all the preparatory and development meetings in the role of specialists in data integration and technology use. Vrailexia is still in progress. All three authors are part of the project team, and one has actively participated in the design of the artifact.

The two projects share many similarities and also have differences. They are equivalent in duration, use the same technologies, and share the same research approach that combines artificial intelligence with virtual reality. They share the objective of developing and deploying an AI and a VR application. Numerous actors with different skills and cultures took part in the processes of both projects and provided their experience and expertise. Related to the differences between these two projects, Schopper is a French ANR project, while Vrailexia is an Erasmus+ European project with 10 partners. Schopper’s end users are archeologists, while Vrailexia’s end users are university students with dyslexia. Therefore, the comparison between the two projects is interesting because it considers their similarities and differences. Both projects are related to the decision-making process, which makes the theoretical framework of Shrestha et al. (2019) suitable as an instrumental theory.

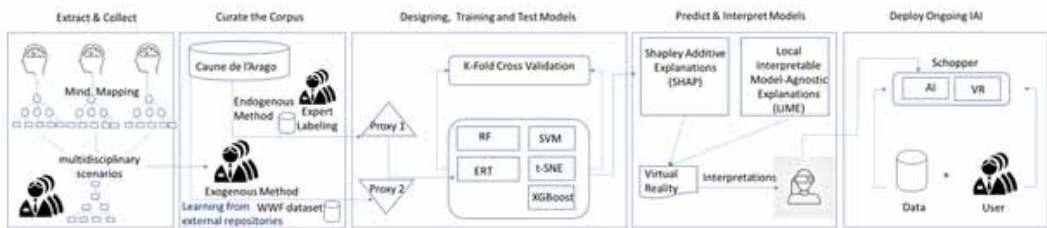
## RESEARCH RESULTS

In this section we present the key phases of the two designs underlying how, in each project, the biases we consider can be reduced or solved by the introduction of a strict collaboration between the experts of the domains and the AI systems.

### Schopper Results

Since project Schopper is finished, all the stages of the design were completed and their contributions and roles in solving the AI biases were superbly detailed. Figure 4 illustrates Schopper’s design. The major stages of the design were done by what we termed an “expert duo.” This duo was composed of the main data scientist and the archeologist who worked on the project full time.

Figure 4. Schopper’s design



### *Extract and Collect*

At the beginning of the project we assumed the presence of metadata related to an integrated database, but when the experts proposed their different scenarios, we discovered that this was not the case. Therefore, the first stage was to perform a knowledge extraction, or mind mapping, and integration to improve our understanding of the data and its context. The research team assisted the archeological experts to design conceptual maps (i.e., multidisciplinary scenarios) with four levels of abstractions: data, facets, constructs, and meta constructs. The first stage focused on the development of a set of scenarios in strict collaboration with the experts.

Regarding the biases related to data collection, the volume of training data was relatively small because it was related to only one part of the excavation efforts. Moreover, there were missing values—some of the target objects were not found or referenced. The techniques to complete missing values (i.e., the minimum number of fragments to be considered instead of presence versus absence of remains) were developed with the help of expert archeologists.

### *Curate the Corpus*

The main objective was to clean the datasets and explore them by applying distinct classical statistical approaches such as correlations, PCA, T-SNE. At the end of this stage, the available data for training and testing were composed of data extracted from the central shared database (Caune de l'Arago) enriched by the many personal databases from the researchers.

For the biases in preprocessing data, the use of statistical approaches allowed the detection of unbalanced classes, points of data with few observations, skewed distributions, and missing values. These problems were fixed with the help of experts in collaboration with the data scientist. Two methods were used to enrich the data: the endogenous (expert labeling) and exogenous (learning from repositories). Indeed, human knowledge was integrated either by the selection of external data or by coding expert opinions.

### *Design*

In the design stage, new designs were used to enrich the data and to choose the suitable algorithms of reference. This was in line with our global approach that included a strict collaboration between the experts and the data scientists.

Datasets were rich in apophenia biases. Matching the correct terminology and identifying when a term was incorrectly used in a certain context was a key point in our analysis. At the same time, the collaboration between experts also helped identify potential false correlations that could have emerged from a simple data-driven analysis.

### *Train and Test*

We used a classic 20/80 split of training data and test data. A close collaboration between the main data scientist and the full-time expert was put in place for the choice of data and algorithms. Several return trips to the previous stages were needed, all decided by the expert duo.

### *Predict and Interpret*

We encountered two different scenarios: 1) when the algorithms were directly interpretable; for example, when the decision model was a tree-based algorithm, the discussion of results was done directly with the experts and 2) when the model was not interpretable, in this case, we use explicability algorithms like SHAP (SHapley Additive exPlanations) to be able to discuss the results with the experts. This stage of the project confirmed the importance of a continuous and effective collaboration between the AI and the experts and users to mitigate the biases linked to transparency and explanation of the results of AI.

### Deploy an Ongoing Augmented Intelligence

This stage began at the end of the Schopper project and did not achieve all of its expected results. The tools developed worked well with the duo of experts but once the project was finished the context became more complicated. The continuous usage by the archaeologists remained extremely targeted and did not benefit from the total integration of the tools.

Following the Shrestha et al. (2019) framework, we identified the following AI use for Schopper with the commented achievements and the discovered opportunities. Table 3 shows that a human-to-AI sequential decision can be an intermediate stage before reaching an aggregated human–AI decision leading to augmented intelligence.

The produced artifact in Schopper’s can be considered as a decision support tool for researchers of archeology. During the development of the project the expert duo collaborated to mitigate biases related to the key stages of the design. This strict and effective collaboration was not foreseen at the beginning. This close team was central in the design of the collaboration between AI and future users. This solved multiple biases related to data collection, data processing, and led to continuous improvements of the results. For example, we can underline the initial difficulties that the data scientist faced to identify the right datasets that could be effectively integrated into each scenario. Any wrong usage of data could have fully corrupted the analysis.

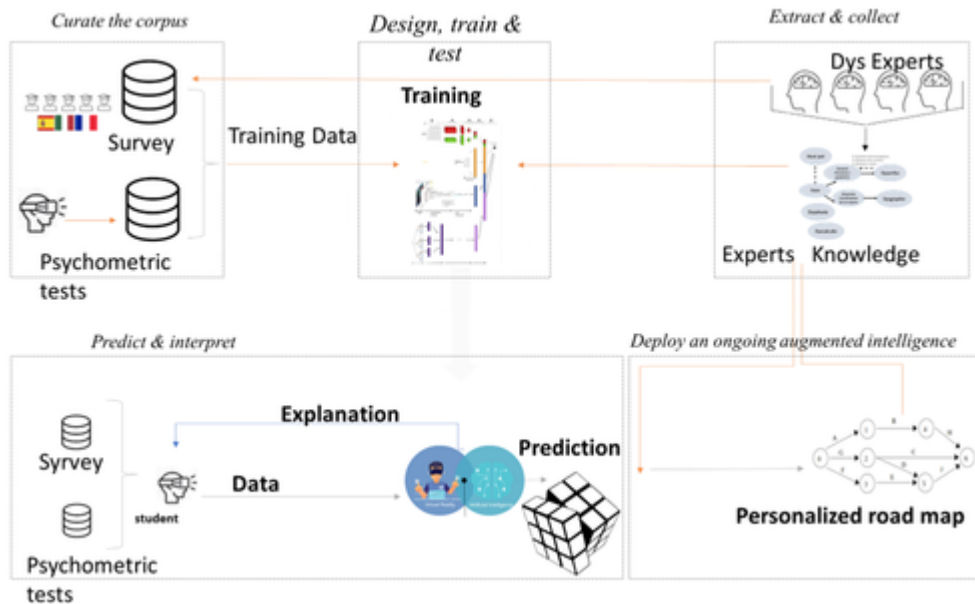
### Vrailexia Results

The Vrailexia project is still under development. This part presents the design and the first results that were achieved during the first three stages (the last stages [train and test, predict and interpret, and deploy an ongoing augmented intelligence] are not detailed in this paper). The preliminary results analyzed in this part show how the guidelines and the experience acquired with Schopper can be reused and applied in a new context. The more promising results come from all the tasks that expect a continuous exchange and collaboration among the users, experts, and AI. Figure 5 depicts the design of Vrailexia.

**Table 3. Types of collaboration between AI and humans in Schopper**

Variables	Schopper
Full Human-to-AI Delegation Not Seen in Schopper	The initial requirement formulated by the archeologists at the beginning of the project was: A full delegation. They requested an AI tool that can predict missions similar to: “Show me the behavior of prehistoric men in a hunting scene.” This would have required a pure connectionist logic which was not possible in this case: “We did not have enough good quality data and no truthful labels for true positive outputs. Since we do not know what was true, the AI couldn’t effectively learn directly from data.”
Hybrid 1: AI-to-Human Sequential Decision Seen but With Group Discussion	The use of SHAP offered explicability for archeologists: “We have had cases of human sequential decision. AI predictions explained by SHAP required group discussions to enrich user visions. This was done without a data scientist.”
Hybrid2: Human-to-AI Sequential Decision Seen but With the Help of Data Scientist	This case is what was called endogenous: “we ask an expert to give his prediction on a series of data and then we explain it by the AI. But it cannot be used without the data scientist. The AI tool is not autonomous.”
Aggregated Human–AI Decision—Augmented Artificial Intelligence	A similar and intuitive example at this level is the treatment of complex questions related to mobility scenarios. This is possible if an integrated platform is developed in future projects. Such a platform requires a more elaborate design and a fully integrated application. Data scientists and experts must follow and collaborate on each phase because it must necessarily include successive iterative improvements.

Figure 5. Vrailexia's design



### *Extract and Collect*

Vrailexia uses data collected from surveys conducted in Spain, France, and Italy. The questionnaire was developed by experts and assessed by dyslexic students in Italy. We did not perform expert knowledge modeling at this stage because the questionnaire was already developed. We realized that we had a lot of data but of poor quality—most of collected data was declarative. We had planned to use medical reports which turned out to be imperfect, and therefore, we could not add factual data. We do not yet have the data from the VR tests because the tests are yet to be completed. The corpus, however, is large and offers opportunities for many statistical and quality analyses that can be reused in the next stages of the project.

In this context, we can identify the classic data biases. Participants who replied to the questionnaires did not form a fully representative sample. Even from a rough analysis, it was clear that some questions were misunderstood by dyslexics. The work done with the experts allowed us to certify and correct some of these biases. For example, there are more females than males in the collected data, while in real life, there are more dyslexic males than females. This bias was discovered thanks to the contribution of an expert. Therefore, experts' contributions were used to filter out the data collected through questionnaire.

### *Curating*

We conducted the first modeling of expert knowledge through more than a dozen interviews and performed advanced statistical analyses on the questionnaire data. On this basis, we have started to clean the data and most likely return to the data collection phase. The comparison of data among countries is complicated because of language differences; for example, Italian is a transparent language, whereas French is an opaque language which changes a lot for dyslexics. The contribution of the experts allowed us to better consider the comorbidities (i.e., links between the different forms of dyslexia) which have strong impacts on learners. These two aspects (language and comorbidities) led to biases in the data that we were able to mitigate based on expert opinion.

### *Design: In Progress*

We have extracted the first model from the data collected by survey using PLS and we are beginning to cross this model with the reasoning taxonomy of the experts, collected by interviews. The use of VR to enrich the data is not yet finalized which prevents us from finishing the design. Therefore, the choice of algorithms is not yet completed, nor is the sequence of treatments.

### *Train and Test, Predict, and Interpret, and Deploy an Ongoing Augmented Intelligence*

These stages are not yet active. Based on Shrestha et al. (2019), Table 4 shows the different potential usages of Vrailexia.

The final result of the project will be a tool to help dyslexic students, it must be autonomous and perform well in hugely different cases. We have sufficient data but they are not of good quality because they are only declarative. We are looking forward to the link with the VR part to get accurate data on individuals. We will need to clean the data and refine the overall models.

## **DISCUSSION AND IMPLICATIONS**

This paper contributes to the Information Systems Theory by opening the black box of designing an aggregation of AI, VR, and humans in organizations and analyzing how it can be done practically. It operationalizes the human-centric AI framework in two design science research (DSR) projects. This research contributes to the existing literature by translating the human-centric AI into a practical process with two complementary parts: 1) a human-in-loop informed design (stages 1 to 5) and 2) an augmented intelligence (stage 6). The two-layer framework emphasizes the importance of integrating human knowledge into AI to mitigate algorithm biases. Table 5 summarizes the different types of knowledge integrated and how they contribute to the biases mitigating in the two projects, Schopper and Vrailexia.

Several lessons can be learned from Schopper while taking into consideration the three ways of working with AI tools:

1. To have substitute AI, the design must be self-sufficient (i.e., there must be enough good quality and tagged data and well-established and stable knowledge rules that allow us to build an autonomous AI artifact that can undertake a process of continuous learning). That was not the case with Schopper.

**Table 4. Types of potential collaborations between AI and humans in Vrailexia**

Variables	Vrailexia
Full Human-to-AI Delegation	We do not think this will be possible or desirable. Each dyslexic student's case is different, and experts are predicting great difficulties in creating categories of dyslexic students. There will be no or very few tools or strategies that work for everyone in every case. It is difficult to have "truth tags" as it was with Schopper.
Hybrid 1: AI-to-Human Sequential Decision	The AI proposes tools, and the student confirms if it suits them. This iteration seems particularly important given the customization of the tools to be proposed.
Hybrid 2: Human-to-AI Sequential Decision	The students perform various tests. Thanks to the results, we can identify with the AI and the VR categories of difficulties that are encountered. In the second step we can associate the difficulties with tools and strategies to overcome them. Finally, the AI analyzes the results and proposes a new cycle of use.
Aggregated Human-AI Decision—Augmented Artificial Intelligence	Be-special, the educational module of Vrailexia, will recommend tools and strategies to the dyslexic student, who will use them, and we will have built-in feedback on their effectiveness. This is the real objective of the project.

Table 5. Stages and steps of human-centric AI to counter AI biases

Human Centric AI	Bias Schopper	Bias Vrailexia	Example of Human Knowledge to Mitigate Biases
Extract and Collect	No real integrated database No available metadata	Sample inadequacy Sample selection bias Out-group homogeneity bias	Resampling (Oversampling and Under sampling) Design of conceptual maps Statistical approaches to detect unbalanced classes, and points of data with few observations
Curate the Corpus	A small volume of data Missing Data	Some questions were misunderstood by dyslexics	New techniques were developed by experts
Design AI	Correlation fallacy Apophenia bias	<i>n.a.</i>	collaboration between experts identified false correlations
Train and Test	Overfitting Underfitting	<i>n.a.</i>	Expert duo K-Fold Cross-Validation and the Confusion matrix
Predict and Interpret	AI algorithms are not transparent, poorly explainable, and not accountable	<i>n.a.</i>	SHAP used by the duo
Deploy Ongoing AI	<i>n.a.</i>	<i>n.a.</i>	<i>n.a.</i>

2. The ongoing collaboration between AI and the end user seems to be an opportunity for continuous checking and improvements. This approach worked well in Schopper and can be classified as an intermediate step between brute substitution and augmentation. The presence of a data scientist is also necessary in various phases.
3. Full collaboration between AI and humans requires integrated tools and a collaborative approach to design such tools. If we separate the expert and the algorithms in the design phase, we cannot bring the users and the algorithms closer together in the final phases.

As Vrailexia was designed based on the experience gained in Schopper, only two lessons can be added:

1. The expert duo was not able to be created as in Schopper. This can be related to the fact that, with Vrailexia, the different partners implicated in the projects are from different countries, which makes such a duo more difficult than within Schopper.
2. We have not yet succeeded in overcoming all the challenges related to the execution of the Vrailexia design. The final global link between AI tools, VR, and end users is not complete. But it is becoming clear that such a design can improve the collaboration between dyslexic students, experts, and AI. This collaboration strongly contributes to solving the apophenia and correlation biases that were already identified in the first phases of the Vrailexia project.

Schopper and Vrailexia show different potential types of delegation and collaborations between AI and humans. The whole delegation requires a fully automated decision without a human implication in its architecture so that everything relates only to data and rules. Such fully automated AI can be designed to solve simple problems. The intermediate level is the succession of AI and humans. It

requires a first level of collaboration in the design at an intermediate stage. Augmented AI requires a collaborative design to prepare the integration of tools and the collaborative use of the application. The major difference between Schopper and Vrailexia is that in the first, the experts are at the same time the end users, while in the second, they are distinct.

Although these two projects have different purposes (i.e., Schopper facilitates the identification of different options and Vrailexia recommends different strategies and tools), it seems there is no change in the treatment of biases. The problems associated with data collection are relatively similar in both cases. With Schopper, however, the biases related to the design and usage could be countered and reduced by a close collaboration between experts and data scientists. This collaboration was possible because experts were at the same time the end users. With Vrailexia, since the experts (i.e., psychologists, researchers, speech therapists, heads of associations) were not end users, collaboration with the data scientists was more difficult.

In past IS research, IT usage has been considered as the use of an inactive artifact by a human actor aiming to reach his objectives. The use has been analyzed by simple models. Our results show that this simplistic view does not fit for advanced artifacts such as AI. New approaches are more suitable. For example, Baird and Maruping (2021) propose the use of an IT artifact as an agency relationship issue (Eisenhardt, 1989) between a principal (the user) and an agent (agentive IS artifact). The principal asks to perform a task that requires a delegation of decisions. The agentive IS artifact refers to rational software-based agents that perceive and act, such as taking on specific rights for task execution and responsibilities for preferred outcomes (Russell, 2019). This model of the agency relationship allows one to consider the context and the aspirations of the user, and the characteristics of the agentive artifact and so precise (to explicit) regarding the relation between user and AI. This approach to the relationship between AI and users gains much of its meaning in our two projects.

## **CONCLUSION**

This research links AI design with AI usage. It fills a gap that was not addressed before. In reference to classical IS theories, we have witnessed a switch from the traditional way of designing AI applications to more collaborative and agile designs. The human-centric AI approach implicates the users in the design and in all stages through implementation and use. We have opened the black box of AI design to explain how organizations can aggregate AI and humans to reach an augmented intelligence.

Our research has several limitations. First, the field consists of only two projects, one of which is still ongoing. The framework can be applied to other cases or projects and a quantitative study can be considered to broaden the scope of the results. Second, we have studied two projects related to decision-making in two different scientific categories (archeology and dyslexia). In both cases, the scientific knowledge associated with the decision-making situation was already well structured. Further research can also explore projects with less structured knowledge to check if the framework is relevant. Third, both cases used do not implicate a time constraint. In cases of high urgency, the consideration of human expertise may be too slow. Future research can help to better analyze wider contexts.



## REFERENCES

- Adam, M. T., Gregor, S., Hevner, A., & Morana, S. (2021). Design science research modes in human-computer interaction projects. *AIS Transactions on Human-Computer Interaction*, 13(1), 1–11. doi:10.17705/1thci.00139
- Akter, S., Dwivedi, Y. K., Sajib, S., Biswas, K., Bandara, R. J., & Michael, K. (2022). Algorithmic bias in machine learning-based marketing models. *Journal of Business Research*, 144, 201–216. doi:10.1016/j.jbusres.2022.01.083
- Baird, A., & Maruping, L. (2021). The next generation of research on IS use: A theoretical framework of delegation to and from argentic IS artifacts. *Management Information Systems Quarterly*, 45(1), 315–341. doi:10.25300/MISQ/2021/15882
- Batista, G., & Monard, M. (2003). An analysis of four missing data treatment methods for supervised learning. *Applied Artificial Intelligence*, 17(5-6), 519–533. doi:10.1080/713827181
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information Communication and Society*, 15(5), 662–679. doi:10.1080/1369118X.2012.678878
- Buhmann, A., & Fieseler, C. (2021). Towards a deliberative framework for responsible innovation in artificial intelligence. *Technology in Society*, 64, 101475. doi:10.1016/j.techsoc.2020.101475
- Camacho, D. M., Collins, K. M., Powers, R. K., Costello, J. C., & Collins, J. J. (2018). Next-generation machine learning for biological networks. *Cell*, 173(7), 1581–1592. doi:10.1016/j.cell.2018.05.015 PMID:29887378
- Chen, J., Lim, C. P., Tan, K. H., Govindan, K., & Kumar, A. (2021). Artificial intelligence-based human-centric decision support framework: An application to predictive maintenance in asset management under pandemic environments. *Annals of Operations Research*. Advance online publication. doi:10.1007/s10479-021-04373-w PMID:34785834
- Collins, C., Dennehy, D., Conboy, K., & Mikalef, P. (2021). Artificial intelligence in information systems research: A systematic literature review and research agenda. *International Journal of Information Management*, 60, 102383. doi:10.1016/j.ijinfomgt.2021.102383
- Crawford, K., & Calo, R. (2016). There is a blind spot in AI research. *Nature*, 538(7625), 311–313. doi:10.1038/538311a PMID:27762391
- Curtarolo, S. G. L., Hart, M. B., Nardelli, N., Mingo, S. S., & Levy, O. (2013). The high-throughput highway to computational materials design. *Nature Materials*, 12(3), 191–201. doi:10.1038/nmat3568 PMID:23422720
- Davison, R. M., Martinsons, M. G., & Ou, C. X. J. (2012). The roles of theory in canonical action research. *Management Information Systems Quarterly*, 36(3), 763–786. doi:10.2307/41703480
- Diligenti, M., Gori, M., & Sacca, C. (2017). Semantic-based regularization for learning and inference. *Artificial Intelligence*, 244, 143–165. doi:10.1016/j.artint.2015.08.011
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10), 78–87. doi:10.1145/2347736.2347755
- Dwivedi, Y. K., Hughes, L., Ismagilova, E., Aarts, G., Coombs, C., Crick, T., Duan, Y., Dwivedi, R., Edwards, J., Eirug, A., Galanos, V., Ilavarasan, P. V., Janssen, M., Jones, P., Kar, A. K., Kizgin, H., Kronemann, B., Lal, B., Lucini, B., & Williams, M. D. (2021). Artificial intelligence (AI): Multidisciplinary perspectives on emerging challenges, opportunities, and agenda for research, practice and policy. *International Journal of Information Management*, 57, 101994. doi:10.1016/j.ijinfomgt.2019.08.002
- Eisenhardt, K. M. (1989). Agency theory: An assessment and review. *Academy of Management Review*, 14(1), 57–74. doi:10.2307/258191
- Faghmous, J. A., Banerjee, S. S., Steinbach, M., Kumar, V., Ganguly, A. R., & Samatova, N. (2014). Theory-guided data science for climate change. *Computer*, 47(11), 74–78. doi:10.1109/MC.2014.335 PMID:25276499
- Grégoire, S., Boulbes, N., Quinio, B., & Boussard, M. (Eds.). (2021). Innovative multidisciplinary method using machine learning to define human behaviors and environments during the Caune de l'Arago (Tautavel, France) Middle Pleistocene occupations. In *Proceedings of UISPP 18th World Congress, Paris, Big Data and Archaeology*. Archaeopress 2021.

- Grønsund, T., & Aanestad, M. (2020). Augmenting the algorithm: Emerging human-in-the-loop work configurations. *The Journal of Strategic Information Systems*, 29(2), 101614. doi:10.1016/j.jsis.2020.101614
- Harfouche, A., Quinio, B., Saba, M., & Bou Saba, P. (2023). The recursive theory of knowledge augmentation (RTKA): Integrating domain knowledge with artificial intelligence to augment organizational knowledge. *Information Systems Frontiers*, 25(3), 55–70. doi:10.1007/s10796-022-10352-8
- Harfouche, A., Quinio, B., Skandrani, S., & Marciniak, R. (2017). *A framework for artificial knowledge creation in organizations*. The 38th International Conference on Information Systems (ICIS2017), Seoul, South Korea.
- Harfouche, A. L., Jacobson, D. A., Kainer, D., Romero, J. C., Harfouche, A. H., Mugnozza, G. S., Moshelion, M., & Tuskan, G. A. (2019). Accelerating climate resilient plant breeding by applying next-generation artificial intelligence. *Trends in Biotechnology*, 37(11), 1217–1235. doi:10.1016/j.tibtech.2019.05.007 PMID:31235329
- Harfouche, A. L., Nakhle, F., Harfouche, A. H., Sardella, O. G., Dart, E., & Jacobson, D. (2022). A primer on artificial intelligence in plant digital phenomics: Embarking on the data to insights journey. *Trends in Plant Science*. doi:10.1016/j.tplants.2022.08.021 PMID:36167648
- Haselton, M. G., Nettle, D., & Andrews, P. W. (2005). The evolution of cognitive bias. In D. M. Buss (Ed.), *The handbook of evolutionary psychology* (pp. 724–746). John Wiley & Sons Inc.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *Management Information Systems Quarterly*, 28(1), 75–105. doi:10.2307/25148625
- Hinton, G., & Salakhutdinov, R. (2006). Reducing the dimensionality of data with neural. *Science*, 313(5786), 504–507. doi:10.1126/science.1127647 PMID:16873662
- Hirschheim, R. (1985). User experience with and assessment of participative systems design. *Management Information Systems Quarterly*, 9(4), 295–303. doi:10.2307/249230
- Horvatić, D., & Lipić, T. (2021). Human-centric AI: The symbiosis of human and artificial intelligence. *Entropy (Basel, Switzerland)*, 23(332), 332. doi:10.3390/e23030332 PMID:33799841
- Huang, M. H., & Rust, R. T. (2021). A strategic framework for artificial intelligence in marketing. *Journal of the Academy of Marketing Science*, 49(1), 30–50. doi:10.1007/s11747-020-00749-9
- Issa, T., & Isaías, P. (2022). *Sustainable design: HCI, usability and environmental concerns* (2nd ed.). Springer., doi:10.1007/978-1-4471-7513-1
- Jain, H., Padmanabhan, B., Pavlou, P. A., & Santanam, R. T. (2018). Call for Papers - Special Issue of Information Systems Research - Humans, Algorithms, and Augmented Intelligence: The Future of Work, Organizations, and Society. *Information Systems Research*, 29(1), 250–251. doi:10.1287/isre.2018.0784
- Johnson, M., AlBizri, A., Harfouche, A., & Fosso-Wamba, S. (2022). Integrating human domain knowledge into artificial intelligence: Informed artificial intelligence. *International Journal of Information Management*, 64, 102479. doi:10.1016/j.ijinfomgt.2022.102479
- Johnson, M., AlBizri, A., Harfouche, A., & Tutun, S. (2021). Digital transformation to mitigate emergency situations: Increasing opioid overdose survival rates through explainable artificial intelligence. *Industrial Management & Data Systems*, 123(1), 324–344. doi:10.1108/IMDS-04-2021-0248
- Kaplan, A. M., & Haenlein, M. (2019). Digital transformation and disruption: On big data, blockchain, artificial intelligence, and other things. *Business Horizons*, 62(6), 679–681. doi:10.1016/j.bushor.2019.07.001
- Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., Shekhar, S., Samatova, N., & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10), 2318–2331. doi:10.1109/TKDE.2017.2720168
- Lundberg, S. M., & Lee, S. I. (2017). *A unified approach to interpreting model predictions* [Paper presentation]. The 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, United States.
- Luo, Y., Cuneo, K. C., Lawrence, T. S., Matuszak, M. M., Dawson, L. A., Niraula, D., Ten Haken, R. K., & El Naqa, I. (2022). A human-in-the-loop based Bayesian network approach to improve imbalanced radiation outcomes prediction for hepatocellular cancer patients with stereotactic body radiotherapy. *Frontiers in Oncology*, 9(12), 1061024. doi:10.3389/fonc.2022.1061024 PMID:36568208

- Marnewick, C., & Marnewick, A. L. (2020). The demands of industry 4.0 on project teams. *IEEE Transactions on Engineering Management*, 67(3), 941–949.
- Martin, K. (2018). Ethical implications and accountability of algorithms. *Journal of Business Ethics*, 160(4), 835–850. doi:10.1007/s10551-018-3921-3
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2). doi:10.1177/2053951716679679
- Nahavandi, S. (2019). Industry 5.0—A human-centric solution. *Sustainability (Basel)*, 11(16), 4371. doi:10.3390/su11164371
- Pecorelli, F., Di Nucci, D., De Roover, C., & De Lucia, A. (2020). A large empirical assessment of the role of data balancing in machine-learning-based code smell detection. *Journal of Systems and Software*, 169, 110693. doi:10.1016/j.jss.2020.110693
- Potelle, H., & Leblond, L. (2018). How to succeed in data science projects industrialization? *Management and Data Science*, 3(1). doi:10.36863/mds.a.4717
- Quinio, B., Boulbes, N., De Pechpeyrou, P., & Kotras, B. (Eds.). (2020). Use cases of virtual reality to visualize a database - How useful is VR for archaeology researchers? In *Proceedings of the Digital Tools & Uses Congress (DTUC'20), October 15–17, 2020, Hammamet, Tunisia*. ACM.
- Rai, A., Constantinides, P., & Sarker, S. (2019). Next-generation digital platforms: Toward human–AI hybrids. *Management Information Systems Quarterly*, 43(1), iii–ix.
- Robert, A. D., & Bouillaguet, A. (1997). *L'analyse de contenu. Que sais-je?* PUF.
- Rokach, L. (2010). Ensemble-based classifiers. *Artificial Intelligence Review*, 33(1), 1–39. doi:10.1007/s10462-009-9124-7
- Rožanec, J. M., Novalija, I., Zajec, P., Kenda, K., Ghinani, H. T., Suh, S., Veliou, E., Papamartzivanos, D., Giannetos, T., Menesidou, S. A., Alonso, R., Cauli, N., Meloni, A., Recupero, D. R., Kyriazis, D., Sofianidis, G., Theodoropoulos, S., Fortuna, B., Mladenicić, D., & Soldatos, J. (2022). Human-centric artificial intelligence architecture for industry 5.0 applications. *International Journal of Production Research*, 61(20), 6847–6872. doi:10.1080/00207543.2022.2138611
- Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control*. Viking Press.
- Shrestha, Y. R., Ben-Menahem, S. M., & Von Krogh, G. (2019). Organizational decision-making structures in the age of artificial intelligence. *California Management Review*, 61(4), 66–83. doi:10.1177/0008125619862257
- Smith, A. (2018). Franken-algorithms: the deadly consequences of unpredictable code. *The Guardian*. <https://www.theguardian.com/technology/2018/aug/29/coding-algorithms-frankenalgos-program-danger>
- Tsamados, A., Aggarwal, N., Cowls, J., Morley, J., Roberts, H., Taddeo, M., & Floridi, L. (2021). The ethics of algorithms: Key problems and solutions. *AI & Society*, 2021(1), 1–16. doi:10.1007/s00146-021-01154-8
- Tutun, S., Harfouche, A., Albizri, A., Johnson, M., & Haiyue, H. (2022). A responsible AI framework to mitigate the ramifications of organ donation crisis. *Information Systems Frontiers*. Advance online publication. doi:10.1007/s10796-022-10340-y
- Von Rueden, L., Mayer, S., Beckh, K., Georgiev, B., Giesselbach, S., Heese, R., Kirsch, B., Pfrommer, J., Pick, A., Ramamurthy, R., Walczak, M., Garcke, J., Bauckhage, C., & Schuecker, J. (2023). Informed machine learning - Towards a taxonomy of explicit integration of knowledge into machine learning. *IEEE Transactions on Knowledge and Data Engineering*, 35(1), 614–633. doi:10.1109/TKDE.2021.3079836
- Wilson, H. J., & Daugherty, P. R. (2018). Collaborative intelligence: Humans and AI are joining forces. *Harvard Business Review*, 96(4), 114–123.
- Yu, K., Beam, H. A. L., & Kohane, I. S. (2018). Artificial intelligence in healthcare. *Nature Biomedical Engineering*, 2(10), 719–731. doi:10.1038/s41551-018-0305-z PMID:31015651
- Ziewitz, M. (2015). Governing algorithms: Myth, mess, and methods. *Science, Technology & Human Values*, 41(1), 3–16. doi:10.1177/0162243915608948 PMID:25866425

## APPENDIX A

### Schopper

#### *Test and Dataset*

Six main tests were performed, each following this cycle:

- Choice of the question or scenario by the archaeologists to be treated.
- Constitution of the dataset.
- Collaborative work between the data scientist and the archaeologist in charge of the tests (this could take several weeks with different exchanges and backtracking on the question or the dataset).
- Internal validation by the data scientist and the archaeologist in charge of the tests.
- Formalization of the Jupiter document.
- Presentation of the results to the archaeologists' team.

Two authors of this paper collected all the datasets and Jupiter documents. They attended, or listened to, all major meetings discussing the test results.

#### *Meeting and Emails*

Since the beginning of the Schopper project in January of 2018 until its conclusion in November of 2020, one of the authors of this paper has participated in all meetings. The meetings covered distinct and were different in nature: a steering committee, work on the AI phase, work on the VR phase, seminar on the project, meeting of tests discussion, etc. Each meeting was reported in detail and the most pivotal were recorded. The documents analysis included about fifty minutes of meeting collaboration, all validated by the project partners. The analysis also included around 2,000 emails exchanged during the project. The six-page word table produced a full timeline of the project since its inception. A chronological presentation can be found at <http://Schopper-anr.org/shopper-wp/29158-2/>

## APPENDIX B

### Vrailexia

#### *Data Collection by Survey*

Data collection by questionnaire was conducted in France, Spain, and Italy. For this article, we used only the French and Spanish data. Table 6 shows the number of questionnaires received and used. The strict application of the RGPD in France obliged us to open the survey to dyslexic and non-dyslexic students, which is not the case for the Spanish answers.

**Table 6. Vrailexia survey**

Country	Non-DYS Answer	DYS Answer	Total
France	1693	301	1994
Spain	0	82	82

#### *Links to the Questionnaires*

<https://vrailexia.eu/survey/>

<https://enquetes.parisnanterre.fr/index.php/389311?lang=fr>

#### *Administration of Questionnaire With Authorization*

<https://enquetes.parisnanterre.fr/index.php/surveyAdministration/view/surveyid/389311>

**Table 7. Question data**

Difficulties
<p>Difficulties in reading</p> <p>Difficulty understanding text</p> <p>Difficulty understanding difficult or unusual words</p> <p>Difficulty understanding lessons</p> <p>Difficulty concentrating during individual study</p> <p>Difficulty paying attention during face-to-face classes</p> <p>Difficulty paying attention during online courses</p> <p>Difficulty remembering concepts just studied</p> <p>Difficulty remembering during review of concepts studied</p> <p>Difficulty organizing time and studying</p> <p>Difficulty taking notes</p> <p>Difficulty due to limited time to prepare a task/question/exam</p> <p>Have you encountered any other difficulties not listed above? If so, please indicate which one(s).</p>

*continued on following page*

**Table 7. Continued**

<b>Tools for Study</b>
<p>Audio book with human voice  Audio book with robotic voice  Words written in different colors  Use of EasyReading font  Use of a pen or smart tablet to take notes and record your voice  Clearer presentation of study material  Having key words in the text already marked  Having concept maps already prepared  Having diagrams ready to go  Having summaries ready to go  E-books (digital books)  Digital tutor (like Siri) to ask for unclear concepts  Images to help understand and remember difficult single words  Images to help memorize and retain a concept  Audio recording of lessons  Video lessons  Have the ability to integrate study materials with internet searches</p>
<b>Study Assistance Strategy</b>
<p>Having someone read for you  Self-made concept maps  Self-made diagrams  Self-made summaries  Repeat the content you have studied  Highlighting key words  Underline the texts with different colors  Be part of a study group  Have a tutor  Create a dyslexic student association to exchange information and resources  Attend face-to-face lessons  Have online lessons  Take breaks during lessons  Have slides of lessons available  Recording lessons  Take notes  Have the course syllabus available in advance  Having the ability to divide an exam, homework, oral quizzes into parts  Having fully written exams or tests  Have fully oral exams or tests  Take exams, assignments, and oral quizzes with only the teacher present  Have an online database with notes, diagrams, summaries, etc. made by other students  Do you know of any other support tools or strategies that are not on the previous two lists that you think would be helpful? If so, please indicate which one(s).</p>

### ***Data Collection by Interviews***

Two different researchers conducted ten interviews with dyslexia experts (psychologists, neurologists, speech therapists, and teachers). They were semi-structured interviews that were fully transcribed by the interviewer. Sixty pages of verbatim transcripts were analyzed using Nvivo.

### ***The Guideline for Interviews***

**Researcher (R0):** Short presentation of the Vrailexia project, authorization to record and presentation of the expert and his or her link with Dyslexia.

- R1:** What are the essential characteristics of dyslexic learners, or DYS more generally, that we need to consider in order to help them?
- R2:** I understand that a dyslexic may make significant errors on one part of a sentence and not at all on another. How can this be explained?
- R3:** I understood that there are two types of DYS difficulties: where one does badly and where one can't complete a task at all. Is that right?
- R4:** In the project we are going to consider the soft skills such as self-confidence, does that seem useful to you?
- R5:** In the personal and family environment, what do you think are the key elements to consider for dyslexic students?
- R6:** I understand that outside noise, and more generally the environment, can have an especially important impact on the activity of dyslexic students. Do you agree?
- R7:** What are the best ways to limit distraction or increase concentration for dyslexics (exam situation or others)?
- R8:** Which other DYS (e.g., dyspraxia and dyscalculia) do you think are the most common with dyslexia?
- R9:** How important are the structures for DYS at the University and for the training of teachers?
- R10:** If we improve pedagogy, we will improve teaching for DYS students. Can we say that we will improve teaching for all students?
- R11:** Which of the DYS are often associated with, or on the contrary, stand out from the others? Are there any general rules or are they just special cases?
- R12:** Are some studies impossible or exceedingly difficult for some DYS?
- R13:** Is there a difference between boys and girls who have DYS?

*Antoine Harfouche is an Associate Professor of Information Systems (IS) and Artificial Intelligence (AI) at the University Paris Nanterre. He has contributed to the IS and AI communities through his excellent teaching, innovative research, and outstanding service. He was awarded the AIS Sandra Slaughter Outstanding Service Award in 2020. Dr. Harfouche completed his MS and PhD in Management IS at Paris Dauphine University (in collaboration with Georgia State University). His research focuses primarily on how IS and AI impact organizations and societies in general. His publications have appeared in peer-reviewed journals such as the International Journal of Information Management, International Journal of Production Research, Industrial Management & Data Systems, the Annals of Operation Research, Information Systems Frontiers, Trends in Plant Science, Trends in Biotechnology, Information Technology and People, Journal of Global Information Management, Journal of Organizational and End User Computing, Research and Applications in Marketing, and Lecture Notes in Information Systems and Organization. He is also a member of the editorial advisory board of the Journal of Enterprise Information Management and is Associate Editor of the Journal of Decision Systems.*

*Bernard Quinio is an Assistant Professor in management at the University Paris Nanterre. An engineer by training and holder of a DEA in applied mathematics, after a professional experience in the private sector, he embarked on a doctorate in management at the University of Nantes. He was deputy director of the UFR Segmi, then, in 2012, Vice president of Paris Nanterre University in charge of continuing education and relations with the territory. In this context, he dealt with entrepreneurship, professional integration and relations with companies. He left the vice-presidency to take charge of the Continuing Education Service in 2017. He was responsible for UPN of an ANR project on the use of artificial intelligence and virtual reality for archeology (ANR Schopper from 2017 to 2020), and now he is responsible of an Erasmus + project on dyslexia (since 2020). He currently teaches in masters in the field of information systems and technologies' usages.*

*Francesca Bugiotti works at CentraleSupélec as an assistant professor in the computer science department. She is member of the équipe Lahdak (LISN) and focuses her research on NoSQL storage systems integration, investigating NoSQL data model characteristics, and query expressive power. She received her PhD in Computer Science from Università Roma Tre. She authored her thesis on the heterogeneity in databases under the supervision of Professor Paolo Atzeni. During her PhD work, she also interned at Inria Saclay, studying the problem of indexing RDF datasets in a cloud infrastructure.*