AI-based User Empowering Use Cases for Visual Big Data Analysis

Thoralf Reis[a], Sebastian Bruchhaus[a], Florian Freund[a], Marco X. Bornschlegl[a] and Matthias L. Hemmje[a]

[a]*University of Hagen, Faculty of Mathematics and Computer Science, 58097 Hagen, Germany*

Abstract

This paper aims to assess scenarios where AI empowers experts and end users carrying out visual Big Data Analyses by systematically deriving and classifying use cases that interconnect these trending research and application areas. Thereby, and to utilize a unified terminology, the results align with the AI2VIS4BigData Reference Model and its Service-Oriented Architecture, a conceptual framework for visual Big Data Analysis in combination with AI. The modeling of use cases with practical relevance within this paper follows a two step approach: An existing reference model for visual Big Data Analysis is reviewed by conducting a cognitive walk-through and the revealed challenges are utilized to define a set of use cases that drive existing research forward. These use cases are subsequently validated utilizing the result of an exploratory survey by the aid of a group of international scientists.

Keywords

User Empowerment, AI, Big Data, Visualization, Big Data Analysis, AI2VIS4BigData, Use Cases

1. Introduction and Motivation

The concept of user empowerment in **Information Systems (IS)** comprises methods and principles that aim at increasing the users' motivation and self-confidence to utilize as many of the system's capabilities as possible [1]. Thereby, the users can tailor the interface or adapt the usage of the IS to be more goal-oriented and efficient. Kim et al. summarize IS user empowerment to be strongly related to four psychological aspects: the *"individual's belief in his or her capability to use the system"* [1], a clear understanding and prioritization of the system activities, the awareness which decisions can be made to influence the IS, and the knowledge of the *"degree to which an individual can influence task outcomes based on the use of system"* [1].

Artificial Intelligence (AI) is a collective term for methods and techniques like, e.g., **Machine Learning (ML)** [2] and becomes more and more relevant for practical applications in health care, driverless cars, or humanoid robots [3]. AI and Big Data Analysis are closely connected to each other [4] as Big Data Analysis enables deriving, validating, applying, and improving AI models while AI-driven algorithms support the exploration of Big Data [4].

Although Big Data is popular in both science and industry, definitions of its terminology remain rather unclear. From a global perspective, Big Data *"refers to the explosion of available information"* [5]. A more formal definition can be derived from Doug Laney's data management challenges [6]

CERC 2021: Collaborative European Research Conference, September 09–10, 2021, Cork, Ireland
thoralf.reis@fernuni-hagen.de (T. Reis); sebastian.bruchhaus@fernuni-hagen.de (S. Bruchhaus); florian.freund@fernuni-hagen.de (F. Freund); marco-xaver.bornschlegl@fernuni-hagen.de (M.X. Bornschlegl); matthias.hemmje@fernuni-hagen.de (M.L. Hemmje)
0000-0003-1100-2645 (T. Reis); 0000-0002-7783-2636 (S. Bruchhaus); 0000-0002-7344-6869 (F. Freund); 0000-0003-3789-5285 (M.X. Bornschlegl); 0000-0001-8293-2802 (M.L. Hemmje)

which comprise three dimensions (the three v's); variety (*"high dimensionality"* [5], ambiguous data manifestations), volume (*"massive sample size"* [5]), and velocity (high data rates) [6].

Since *"human brain tends to find pattern more efficiently when data is represented visually"* [7], visualization is an important link between AI and Big Data Analysis [8] that *"enhances comprehension and decreases entry barriers for new users"* [9]. Its purpose ranges from simply presenting information, confirming assumptions to exploring new insights [8].

In 2020, the authors of this paper introduced AI2VIS4BigData, a reference model for scientific and industrial applications that apply AI, Big Data Analysis, and visualization [4]. The reference model's objective is to establish a common terminology, specify relations unambiguously, and enable the derivation [10] for Big Data Analysis systems that utilize visualization and AI [4]. It furthermore presents three use cases that interconnect Big Data Analysis, AI, and visualization. The visualization of these high level use cases in Figure 1 reveals that all use cases are somehow relevant for Big Data Analysis as well as AI user stereotypes. However, there exists neither a detailed description of these use cases nor an assessment to which extent they support user empowerment.

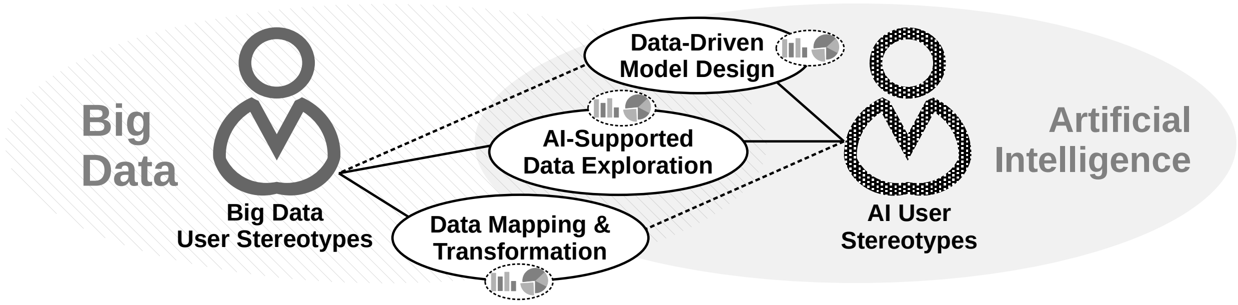

Figure 1: Use Cases interconnecting AI, Big Data Analysis, and Visualization [4]

In this way, this paper intends to provide both detailed descriptions and a classification of use cases with special focus on user empowerment. This shall be accomplished through specifying the use cases and creating a use case taxonomy. For achieving this objective, this paper follows the research approach introduced in [11] to assess manual activities that exist for visual Big Data Analysis over all AI2VIS4BigData processing steps from data integration to data analysis, data visualization, and data exploration as well as review existing challenges in literature to examine the three use cases from Figure 1 in more detail. All manual activities and challenges will then be utilized to derive use cases that solve the underlying problems or mitigate the negative impact of the challenges. Thereby, these use cases empower the users. Finally, these use cases will be clustered and hierarchically categorized within a use case taxonomy. The remainder of this paper contains a presentation of the state of the art (Section 2) with a detailed introduction of the AI2VIS4BigData Reference Model as well as a model review of it based on a cognitive walk-through. Section 3 introduces a derivation of detailed use cases as well as their relationships within a use case taxonomy. Section 4 validates the use cases initially before Section 5 summarizes the results and outlines future research directions.

2. State of the Art

This section outlines selected state of the art with relation to user empowerment and lays the foundation to identify potential use cases of AI for user empowerment to improve visual Big Data Analysis in the AI2VIS4BigData Reference Model through conducting an cognitive walk-through based model review. The cognitive walk-through highlights potential problems as well as manual activities that can consume a significant amount of time.

2.1. AI2VIS4BigData Reference Model

AI2VIS4BigData Reference Model [4] is founded on Bornschlegl's IVIS4BigData Reference Model for visual Big Data Analysis [12]. It extends the visual Big Data Analysis reference model on all aspects of AI and machine learning for data analytics. Its objective is to establish a common terminology, specify relations unambiguously, and enable the derivation [10] for Big Data Analysis systems that utilize visualization and AI [4]. This reference model has been successfully validated in an expert survey as well as an expert workshop with scientists and researchers from six different countries [13]. Further work includes the introduction of a conceptual architecture that is based on this reference model [9]. This paper addresses two shortcomings in the state of the art: Use cases for user empowerment have not been analyzed methodically in the context of Big Data visualization and AI so far. Furthermore the concept of user empowerment in AI2VIS4BigData itself is not comprehensively described as of yet. The reference model is visualized in Figure 2.

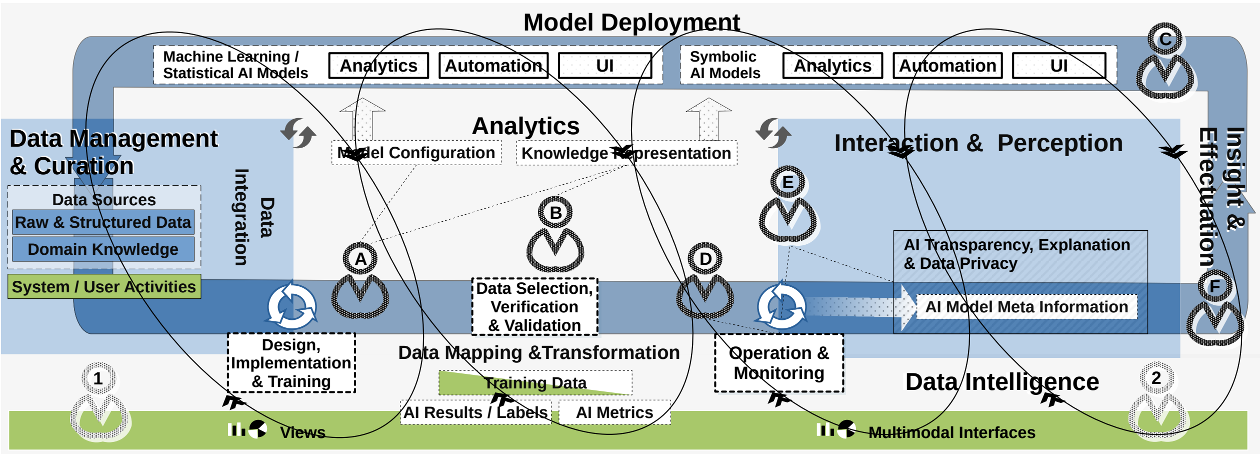

Figure 2: AI2VIS4BigData Reference Model [4] for AI-based support of visual Big Data Analysis

Value generation in AI2VIS4BigData is separated into three dimensions; firstly the transformation of raw data into visualizations of analyzed data, and secondly the transformation from data into information, knowledge, and wisdom [14]. The first dimension is represented through the four consecutive process stages of AI2VIS4BigData [4]. The second dimension is represented through a data intelligence layer that interconnects all process stages and enables the different expert and end user stereotypes that are involved in visual Big Data Analysis to interact with the system and its processing results [14]. Within this reference model, *"end users know the application domain"* [14] while expert users *"are able to configure technical details"* [12]. In regards to end users, it empowers them *"to configure, simulate, optimize, and run each phase of the IVIS pipeline in an interactive way"* [14]. The third dimension is the application of AI within the model deployment layer that lays the foundation for AI-based user empowerment [4]. As visualized in Figure 2, the view on each processing step, the processing step itself, and the processed artifact are interconnected with the respective end user stereotype through four SECI[1] cycles. These cycles represent the consecutive transformation of *"explicit knowledge to tacit knowledge"* [14] and the other way round in order to generate novel knowledge.

2.2. Fischer and Nakakoji's Multifaceted Architecture

User empowerment in context of AI2VIS4BigData comprises user interface configurability, simulations, continuous improvements, and knowledge creation from IVIS4BigData [14]. Hence it is closely related [12] with Fischer and Nakakoji's multifaceted architecture [15]. This architecture consists five elements sorted into three layers [15]: a domain knowledge layer that contains knowledge in form of a catalog of rules and patterns, their meaning for the current application domain as well as the users' comprehension of their significance [15]; a design creation layer which contains a specification, a description of the required characteristics of a design, and a construction, the actual implementation of the design [15]; a feedback layer connects domain knowledge and design creation through critical reflections, case-based reasoning, and simulation [15]. Instead of trying to create an expert system for a well-understood problem that can be fully specified, Fischer and Nakakoji propose to utilize this domain knowledge to empower users to solve the problem by themselves [15]. Their architecture enables end user with lower expertise in the application domain to benefit from the existing knowledge base whereas expert users like, e.g., *"experienced designers"* [15] are empowered to increase the collection within this knowledge base [15].

2.3. Discussion and Remaining Challenges

AI2VIS4BigData and its foundation IVIS4BigData incorporate principles like continuous improvements and the utilization of knowledge [14] to improve the system's usability in a manual form depending on knowledge input and system adaptions by expert users. An application of Fischer and Nakakoji's multifaceted architecture and its feedback layer to empower users based on a knowledge base without requiring expert users to manually adapt the system's user interface remains a challenge. This paper targets to address this challenge by systematically deriving use cases for applications of Fischer and Nakakoji's multifaceted architecture on AI2VIS4BigData.

3. Conceptual Modeling

Section 2 introduced the AI2VIS4BigData Reference Model, Fischer and Nakakoji's multifaceted architecture, and emphasized on combining their underlying concepts. This section reviews the manual activities within AI2VIS4BigData for the purpose of visual Big Data Analysis in order to reveal manual activities and content-related problems which a consequent application of Fischer and Nakakoji's multifaceted architecture can address. These weak spots then serve as basis for conceptually modeling specific use cases based onthe three high-level use cases from Figure 1. The objective is an overview as well as a taxonomy of user empowering use cases.

In order to establish a clear terminology in regards to the application areas' relationship like, e.g., is AI applied for the purpose of visual Big Data Analysis or vice versa, the use cases shown in Figure 1 are renamed: Data-driven model design is renamed to Big Data-supported AI model design as Big Data Analyses are applied for the purpose of designing an AI model; data mapping and transformation is redefined to AI-based Big Data transformation since AI is applied to extract relevant information

[1] Socialization, Externalization, Combination, and Internalization

from Big Data; and in order to emphasize on the importance of user empowerment, AI-supported data exploration is renamed to AI-based Big Data user empowerment.

3.1. Model Review based on a Cognitive Walk-Through

Fischer and Nakakoji's user empowering architecture targets to equip users with relevant insights and knowledge to solve the problem themselves instead of trying to utilize AI to replace them [15]. They mention the challenges of *"inform[ing] and support[ing] the judgment"* [15] of the user as well as of *"automat[ing] tasks that people consider tedious or uninteresting"* [15] to be crucial. To properly support the users, it is crucial to be aware of potential problem types that can occur during visual Big Data Analysis. Consequently, this model review looking into manual, repetitive activities that need to be carried out by human experts and end users looks like a promising starting point for deriving valid user empowerment use cases as well as typical problems that occur thereby.

The model review of AI2VIS4BigData is structured alongside the four Big Data processing steps of its foundation IVIS4BigData since they comprise all Big Data Analysis related activities that can be supported through the application of AI. Beginning with data integration, the expert and end users are required to manually configure the system to adapt to certain data models, data schema, as well as semantics [12]. They then have to configure appropriate wrapper and mediator components before they can simulate and perform the actual data integration [12]. Relevant problems that can thereby occur comprise data level inconsistencies and a high heterogeneity [16] as well as the handling of the enormous amount high-dimensional data [5] which leads to *"scalability"* [5] issues as well as *"storage bottleneck[s]"* [5].

During the data transformation processing step, the users configure and select an appropriate Big Data analysis method as well the data they want to transform [12]. Finally, they configure, simulate, and finally execute the workflow [12]. Relevant problems in this context result on the one hand from bad data quality: *"noise, outliers, low precision, missing values"* [8], errors in measurement [8, 5], duplicates [8] and *"not maintained attribute[s]"* [17]; On the other hand, the massive amounts of data and the dynamic nature of Big Data with *"streams of time related or real time data"* [8] challenge data transformation significantly [5] since *"linear pass of the whole dataset [is] unaffordable"* [5] and these huge data amounts call for *"immense parallelization"* [7].

The third processing step, visual mapping, consists of configuring and selection activities for a visual representation and a visualization library [12]. Selecting the data that shall be visualized, configuring and simulating the visualization workflow conclude the manual visual mapping activities [12]. Problems for this processing step are closely linked to data quality since *"visual noise"* [7], *"spurious correlation, incidental endogeneity"* [5], or *"incidental homogeneity"* [5] make comprehensible visualizations hard to achieve. This applies especially, as the human receptivity is challenged and often overstrained by Big Data's data amount and data velocity [7].

The last processing step view transformation comprises the configuration of IVIS techniques, the selection of visualizations as well as the configuration and simulation of views [12]. All derived manual activities as well as relevant problems in visual Big Data Analysis per AI2VIS4BigData transformation are summarized as results of this cognitive walk-through in Table 1.

3.2. AI-based Big Data Transformation

The AI2VIS4BigData SOA presented in [9] contains AI-based services for every visual Big Data Analysis processing step from data integration to data exploration. Figure 3 visualizes these services together with the introduced SECI cycles and associates them with the use case AI-based Big Data Trans-

Table 1
Manual Activities and Problems in Visual Big Data Analysis derived from an AI2VIS4BigData [4] Model Review

Big Data Transformation	Manual Activity	Problem
Data Integration	Wrapper Configuration Mediator Configuration Data Schema Configuration Data Model Configuration Semantic Resource Configuration Semantic Resource Selection Data Integration Configuration and Simulation	Inconsistency in Data Levels [16] High Data Rate [5]
Data Transformation	Analysis Method Configuration Analysis Method Selection Raw Data Selection Analysis Method Workflow Configuration and Simulation	High Data Dimensionality [5] Measurement Errors & Noise [8, 5] Outliers [8] Missing Values [8, 5] Duplicate Records [8, 17]
Visual Mapping	Visual Representation Configuration Visual Representation Selection Visualization Library Configuration Visualization Library Selecton Structured Data Selection Visualization Workflow Configuration and Simulation	Visual Noise [7] Spurious Correlation [5] Incidental Endogeneity & Homogeneity [5]
View Transformation	IVIS Technique Configuration Visualization Selection View Configuration and Simulation	

formation (B). In analogy to the model review (Section 3.1), the four vertical pillars from IVIS4BigData's transformation pipeline [14] can be utilized to detail the use case further on into sub use cases.

Use Case AI-based Data Integration

The first of these sub use cases is *AI-based Data Integration.* Manual activities and present challenges for this transformation step are utilized to detail this use case further on into the following third level use cases: *Wrapper Detection* (manual activity of wrapper configuration), *Mediator Detection* (manual activity of mediator configuration), *Data Schema Detection* (manual activity of data schema configuration and challenge of inconsistency in data levels), *Data Model Detection* (manual activity of data model configuration), *Data Semantics Detection* (manual activities of semantic resource configuration and selection), and *Data Inflow Prediction* (challenge of high data rates).

Figure 4 visualizes the relationship between AI2VIS4BigData and the IVIS4BigData services for data integration: the AI2VIS4BigData data integration services derive metadata and information on semantic representations of the raw data from the data source systems whereas the IVIS4BigData

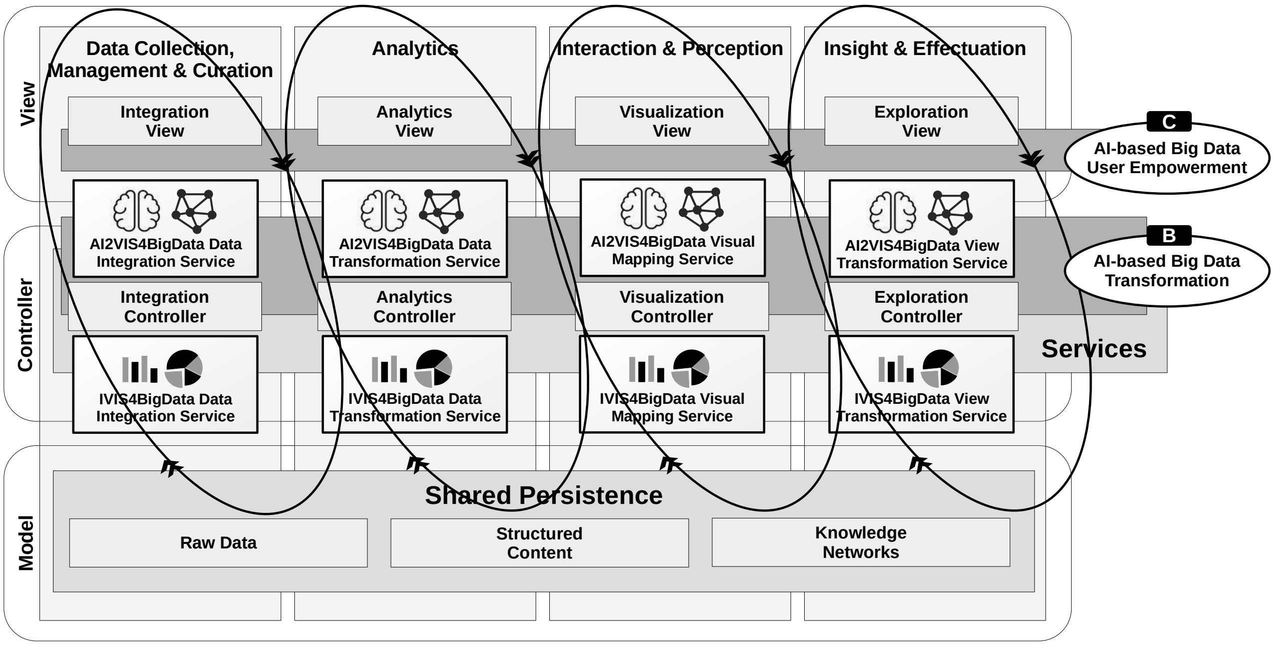

Figure 3: User Empowering Use Cases embedded in the AI2VIS4BigData SOA

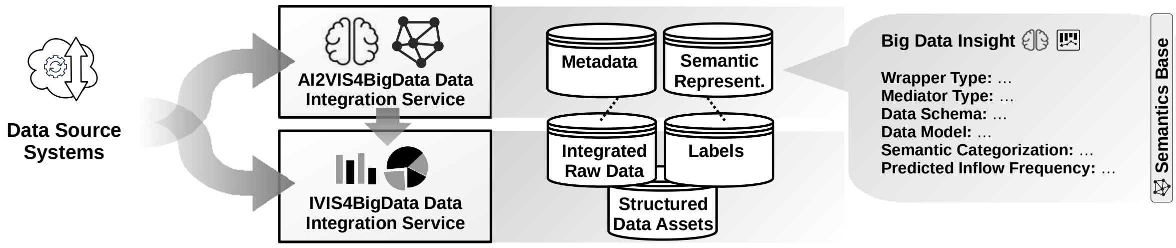

Figure 4: Data Integration Use Case Sketch as Application of AI2VIS4BigData Services

services actually integrate the raw data into the system. The metadata and information on semantic representations are thereby utilized to either assist the human user or to automatize parts of the data integration process.

Use Case B2: AI-based Data Transformation

The data transformation use case for the IVIS4BigData analysis processing phase is visualized as application of an AI2VIS4BigData data transformation service in Figure 5. It consists of the following third level use cases: *Analysis Method Prediction* (manual activities of analysis method configuration and selection), *Raw Data Hotspot Detection* (manual activity of raw data selection, challenge of outliers), *Dimensionality Reduction* (challenge of high data dimensionality, challenge of measurement errors and noise), *Missing Value Detection* (challenge of missing values), and *Duplicate Detection* (challenge of duplicate records).

Figure 5 shows the information flow of the formerly integrated data asses (refer to Figure 4) through both, AI2VIS4BigData and IVIS4BigData services. The different data transformation use cases thereby

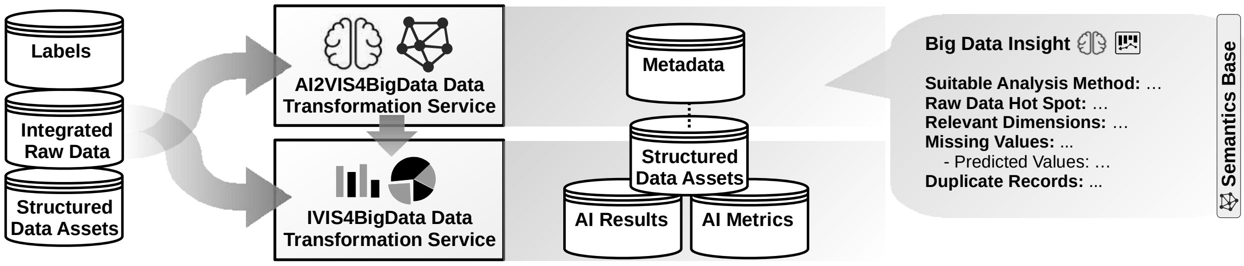

Figure 5: Data Transformation Use Case Sketch as Application of AI2VIS4BigData Services

support the IVIS4BigData transformation through generation metadata.

Use Case B3: AI-based Visual Mapping

The third use case, *AI-based Visual Mapping* (Figure 6), summarizes all application scenarios of AI in order to support the third transformation of IVIS4BigData from analyzed, structured data into visual representations. Consequently, its third level use cases are also derived from manual activities like, e.g, the configuration and selection of suiting visual representations or visualization libraries as well as from challenges like, e.g., visual noise or spurious correlations: *Visual Representation Prediction* (manual activities of visual representation configuration and selection), *Visualization Library Prediction* (manual activities of visualization library configuration and selection), *Structured Data Hotspot Detection* (manual activity of structured data selection and the challenge of incidental endogeneity and homogeneity), *Visual Noise Detection* (challenges of visual noise), and *Correlation Detection* (challenge of spurious correlation).

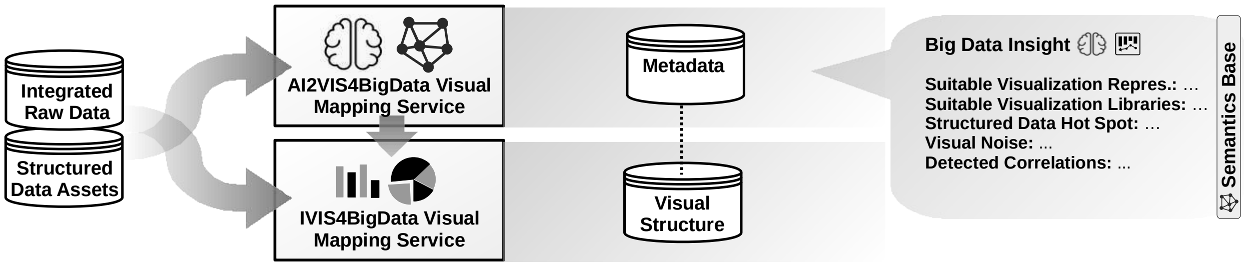

Figure 6: Visual Mapping Use Case Sketch as Application of AI2VIS4BigData Services

Use Case B4: AI-based View Transformation

The fourth and final second level use case derived from IVIS4BigData's data transformation steps is *AI-based View Transformation*. It is visualized in Figure 7 and comprises the transformation of visual structures under application of the selected and configured visualization algorithm into actual views and dashboards that can be perceived by non-expert user stereotypes [12]. Its third level use cases are *IVIS Technique Prediction* (manual activity of IVIS technique configuration), *Visualization Hotspot Detection* (manual activity of visualization selection).

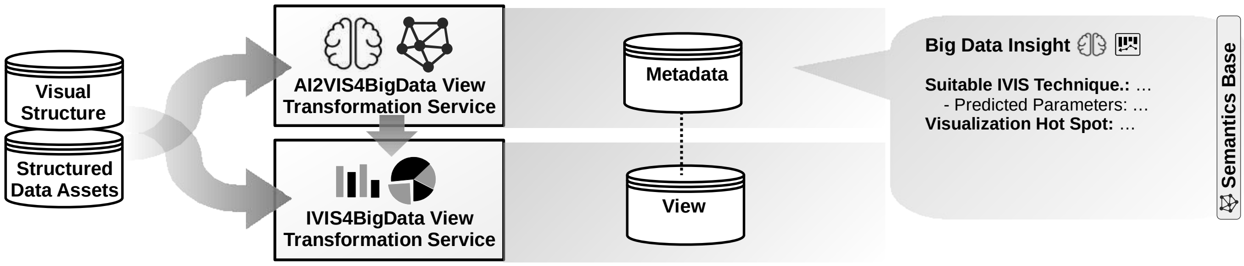

Figure 7: View Transformation Use Case Sketch as Application of AI2VIS4BigData Services

3.3. AI-based Big Data User Empowerment

The authors of this paper propose in [11] that AI-based user empowerment closely follows the principles that Fischer and Nakakoji introduced in their multifaceted architecture [15] with *Design Creation* being the expert and end user stereotype's utilization of the Big Data analysis system with the objective to gain as much insight as possible. Following that interpretation, a *Specification* can be seen as the users' mental model of how he or she thinks that the Big Data exploration objective can be reached, while a *Construction* is the actual articulation of this plan to the system like, e.g., executing a data query or the usage of an UI element. The likelihood that a user decides to follow more sophisticated plans increases with the users self-confidence regarding a holistic system understanding. Positive experiences with successful Big Data Analysis system usage results can be stored as best practices or rules within the *Catalog Base* of the domain knowledge layer while negative experiences can be stored within the *Argumentation Base* of the domain knowledge layer. Expert knowledge regarding Big Data Analysis workflows, useful patterns, and anti-patterns can be integrated in both. The actual meaning of the system and data state can be stored within the *Semantics Base.* The application of case-based reasoning, simulation, and critical challenging can improve the users' self-confidence and thus increase the probability of fruitful Big Data exploration. Figure 8 visualizes the proposed interpretation of Fischer and Nakakoji's multifaceted architecture for visual Big Data Analyses.

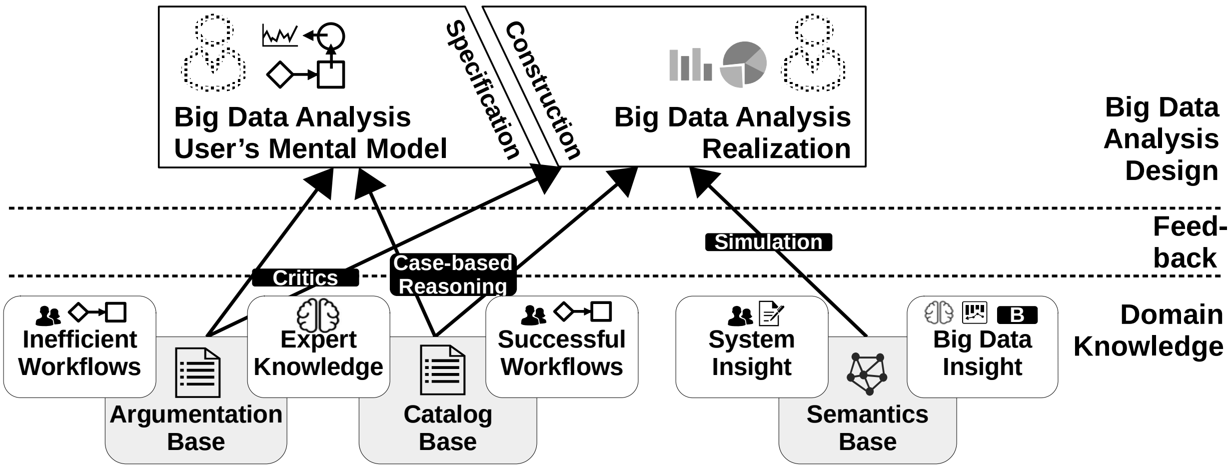

Figure 8: Multifaceted Architecture [15] Interpretation for Visual Big Data Analysis [11]

Following the interpretation in Figure 8, user empowerment for visual Big Data Analysis is closely related to the outcome of use case B: use case B applies AI in order to identify user-relevant information (*Big Data Insight*). Use case (C) then utilizes this insight and applies formalized rules, experiences, or expert knowledge to transport this information to the Big Data Analysis user stereotypes. The form of this information transport is utilized to derive two second-level use cases: *Interaction Guidance* and *Content Guidance.*

These use cases have the objective to address the four psychological principles introduced by Kim et al. [1]: strengthen the individual user's belief in his / her capabilities, sharpen the user's system understanding, raise the awareness which capabilities the system offers, and to clarify the understanding of the users influence on the system [1]. In addition, these use cases aim at minimizing the effect of the presented challenges and thereby follow the **Human-Centered Design (HCD)** approach that intends to improve system usability through *"applying human factors, ergonomics and usability knowledge and techniques"* [18]. According to the ISO9241 standard, the scientific discipline of **Human Factors and Ergonomics (HF/E)** is defined to comprise the research of human interaction with systems as well as all design activities *"to optimize human well-being and overall system performance"* [18]. Hence, use case C and its sub use cases apply methodology and technology from HCD and HF/E.

Use Case Interaction Guidance

Since *"human background knowledge, intuition, and decision-making either cannot be automated or serve as input for the future development of automated processes"* [8], the first use case *Interaction Guidance* focuses on improving the human user's interaction with the system. In more detail, the use case summarizes two different types of guiding the user interaction to make it simpler (addresses the lack of skilled personnel) or more efficient (addresses the challenge of expensive resources). The use case can be implemented as an interaction proposal (e.g. propose the action to fill data sample's missing values if missing values exist or to remove duplicates if duplicate records exist), to prioritize interaction capabilities (e.g. a user interface *"that over time automatically minimizes or even eliminates infrequently used features or menu items"* [19]), or to propose whole workflows (a set of single interactions carried out in a certain order).

Use Case Content Guidance

The second use case *Content Guidance* is closely related to the former one. Its focus lies on informing the user about the content of the Big Data Analysis itself and thereby enable Big Data Analysis user stereotypes to perceive information that might have only been accessible for very skilled and experienced data scientists. Examples range from simple tool-tips (e.g. hints that the current set of data contains missing values or duplicate records) and help dialogues that support the user [16], *"automatic warning messages"* [17] to more complex scenarios such as automatic adaption of visualizations [7] or *"systems that automate the data exploration process by discovering data objects"* [20] steering users *"towards interesting data"* [20].

3.4. Use Case Taxonomy

All derived use cases implicitly created a use case taxonomy through their ordering and through the relationships between each other. This taxonomy is visualized in Table 2. It consists of three use cases on the first level, six on the second level and 18 on the third level.

Table 2
Taxonomy for Use Cases interconnecting AI and Visual Big Data Analysis

Use Cases		
1st Level	2nd Level	3rd Level
Big Data-based AI Model Design		
AI-based Big Data Transformation	AI-based Data Integration	Wrapper Detection
		Mediator Detection
		Data Schema Detection
		Data Model Detection
		Data Semantics Detection
		Data Inflow Prediction
	AI-based Data Transformation	Analysis Method Prediction
		Raw Data Hotspot Detection
		Dimensionality Reduction
		Missing Value Detection
		Duplicate Detection
	AI-based Visual Mapping	Visual Representation Prediction
		Visualization Library Prediction
		Structured Data Hotspot Detection
		Visual Noise Detection
		Correlation Detection
	AI-based View Transformation	IVIS Technique Prediction
		Visualization Hotspot Detection
AI-based Big Data User Empowerment	Content Guidance	
	Interaction Guidance	

Although the formal modeling of all use cases in 2 remains a challenge for future research, a validation of their practical relevance is required. For this purpose, this paper revisits the introduced AI2VIS4BigData expert survey [13] to compare the experts feedback with the derived use cases.

4. Initial Validation of Identified Use Cases

The derived use cases A to C will be initially validated in this Section. For this purpose, two questions from the expert survey in preparation for the AVI 2020 satellite workshop [13] will be reviewed and the experts' answers are associated to the different use cases (upon the second use case level). The result of this initial validation is summarized in Table 3. Validated use cases are visualized with a filled circle (●), non-validated use cases are visualized with an empty circle (○). The following questions have been assessed in the survey:

1. Question 1: What is the practical relevance of a given set of application scenarios for AI, Big Data Analysis, and visualization activities within the survey participant's research? (Table 1 in [13])

2. Question 2: What is the practical relevance of three different AI model types within the survey participant's research? (Figure 10 in [13])

Use case A (*Big Data-based AI Model Design*) is validated through approval of the survey participant's majority for the application scenario *"Applying Big Data to design AI models (e.g. adjust weights of a neural network)"* [13] as well as the high approval rate for the scenario *"Applying visualization to design AI models"* [13].

The survey results for use case B (*AI-based Big Data Transformation*) support only the validation use case the *AI-based Data Integration* (B1) and *AI-based Data Transformation* (B2). B1 is validated through the approval of the workshop participants to *"applying AI to integrate different sources of Big Data"* [13] while use case B2 received the participant's approval for the application scenario *"Applying AI to ease Big Data exploration (e.g. to programmatically identify outliers)"* [13]. The further sub use cases of use case B were not addressed in the survey.

The validation of use case C (*AI-based Big Data User Empowerment*) consists of the combined validation of its sub use cases. Both are validated through the very high approval rate for the application scenario *"Applying AI and Big Data to ease visualization and UI comprehension (e.g. through intelligent UI that explains and highlights useful tools)"* [13] with the explanatory aspect pointing *Content Guidance* and the tool highlighting pointing to *Interaction Guidance*. In addition, the majority of survey participants state at least sometimes to utilize *"UI models"* [13] which strengthens the decision.

Table 3
Summary of Expert Survey [13] based Use Case Validation

Validation Method	Use Case A	Use Case B				Use Case C	
		B1	B2	B3	B4	C1	C2
Expert Survey [13]	●	●	●	○	○	●	●

As Table 3 summarizes, the expert survey [13] validates the use cases A, B1, B2, C1, and C2. A validation of use cases B3 and B4 remains a challenge for future research. Further research directions include a formal UML modeling as well as an extensive validation of all use cases, the introduction of an use case framework, the technical specification of the use cases within the AI2VIS4BigData conceptual architecture (e.g. APIs and data models) as well as a prototypical implementation of them.

5. Conclusion and Outlook

This paper summarizes existing challenges in the application field AI, Big Data Analysis, and visualization, clusters the different user stereotypes involved in these application areas into experts and end users. It introduces 28 AI2VIS4BigData use cases that enable empowering experts and end users involved in the domain of visual Big Data Analysis. Furthermore, these use cases are ordered into a use case taxonomy and validated through analysis of a survey result with answers from international scientists and researchers [13]. Intensive work within one or multiple of the derived use cases, formal modeling, a comprehensive evaluation, a detailed specification, and a proof of concept implementation remain as open challenges and are future research goals.

References

[1] H.-W. Kim, S. Gupta, A user empowerment approach to information systems infusion, IEEE Transactions on Engineering Management 61 (2014) 656–668.

[2] ISO, ISO/IEC JTC 1/SC 42 Artificial Intelligence, 2018. URL: https://isotc.iso.org/livelink/livelink/open/jtc1sc42.

[3] R. Bond, F. Engel, M. Fuchs, M. Hemmje, P. M. Kevitt, M. McTear, M. Mulvenna, P. Walsh, H. J. Zheng, Digital empathy secures Frankenstein's monster, CEUR Workshop Proceedings 2348 (2019) 335–349.

[4] T. Reis, M. X. Bornschlegl, M. L. Hemmje, Towards a Reference Model for Artificial Intelligence Supporting Big Data Analysis, To appear in: Advances in Data Science and Information Engineering -Proceedings of the 2020 International Conference on Data Science (ICDATA'20) (2021).

[5] J. Fan, F. Han, H. Liu, Challenges of big data analysis, National science review 1 (2014) 293–314.

[6] D. Laney, 3D Data Management: Controlling Data Volume, Velocity, and Variety, Technical Report, META Group, 2001.

[7] S. M. Ali, N. Gupta, G. K. Nayak, R. K. Lenka, Big data visualization: Tools and challenges, in: 2016 2nd International Conference on Contemporary Computing and Informatics (IC3I), IEEE, 2016, pp. 656–660.

[8] D. A. Keim, F. Mansmann, J. Schneidewind, H. Ziegler, Challenges in visual data analysis, in: Tenth International Conference on Information Visualisation (IV'06), IEEE, 2006, pp. 9–16.

[9] T. Reis, T. Krause, M. X. Bornschlegl, M. L. Hemmje, A Conceptual Architecture for AI-based Big Data Analysis and Visualization Supporting Metagenomics Research, Proceedings of the Collaborative European Research Conference (CERC) (2020) 264–272.

[10] O. Thomas, Understanding the Term Reference Model in Information Systems Research: History, Literature Analysis and Explanation, Proceedings of the Workshop on Business Process Reference Model (2005) 16–30.

[11] T. Reis, S. Bruchhaus, B. Vu, M. X. Bornschlegl, M. L. Hemmje, Towards Modeling AI-based User Empowerment for Visual Big Data Analysis, Proceedings of the Second Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2021) (2021) 67—-75. URL: http://ceur-ws.org/Vol-2863/{#}paper-07.

[12] M. X. Bornschlegl, K. Berwind, M. L. Hemmje, Modeling end user empowerment in big data applications, 26th International Conference on Software Engineering and Data Engineering, SEDE 2017 (2017) 47–53.

[13] T. Reis, M. X. Bornschlegl, M. L. Hemmje, AI2VIS4BigData: Qualitative Evaluation of a Big Data Analysis, AI, and Visualization Reference Model, To appear in: Lecture Notes in Computer Science LNCS 12585 (2021).

[14] M. X. Bornschlegl, Advanced Visual Interfaces Supporting Distributed Cloud-Based Big Data Analysis, Dissertation, University of Hagen, 2019.

[15] G. Fischer, K. Nakakoji, Beyond the macho approach of artificial intelligence: empower human designers—do not replace them, Knowledge-Based Systems 5 (1992) 15–30.

[16] A. Kadadi, R. Agrawal, C. Nyamful, R. Atiq, Challenges of data integration and interoperability in big data, in: 2014 IEEE international conference on big data (big data), IEEE, 2014, pp. 38–40.

[17] O. Azeroual, Data Wrangling in Database Systems: Purging of Dirty Data, Data 5 (2020) 50.

[18] International Organization for Standardization, ISO 9241–210: 2019 (en) Ergonomics of human-system interaction—Part 210: Human-centred design for interactive systems (2019).

[19] A. Blair-Early, M. Zender, User interface design principles for interaction design, Design Issues 24 (2008) 85–107.

[20] S. Idreos, O. Papaemmanouil, S. Chaudhuri, Overview of data exploration techniques, in: Proceedings of the 2015 ACM SIGMOD International Conference on Management of Data, 2015, pp. 277–281.