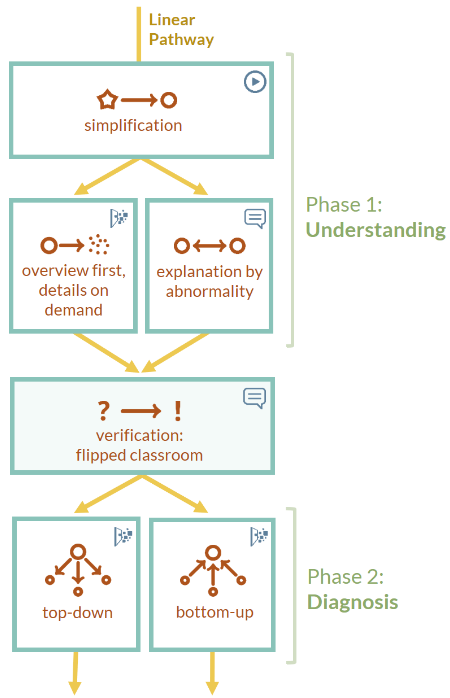

Figure 1: The proposed explanation process model. On the highest level, each explanation consists of different PHASES that structure the whole process into defined elements. Each phase contains EXPLANATION BLOCKS, i.e., self-contained units to explain one phenomenon based on a selected STRATEGY and MEDIUM. At the end of each explanation phase, an optional VERIFICATION BLOCK ensures the understanding of the explained aspects. Lastly, to transition between phases and building blocks, different PATHWAYS are utilized.

Towards XAI: Structuring the Processes of Explanations

Mennatallah El-Assady, Wolfgang Jentner, Rebecca Kehlbeck, Udo Schlegel, Rita Sevastjanova, Fabian Sperrle, Thilo Spinner, Daniel Keim
University of Konstanz
Konstanz, Germany

ABSTRACT

Explainable Artificial Intelligence describes a *process* to reveal the logical propagation of operations that transform a given input to a certain output. In this paper, we investigate the design space of explanation processes based on factors gathered from six research areas, namely, Pedagogy, Storytelling, Argumentation, Programming, Trust-Building, and Gamification. We contribute a conceptual model describing the building blocks of explanation processes, including a comprehensive overview of explanation and verification phases, pathways, mediums, and strategies. We further argue for the importance of studying effective methods of explainable machine learning, and discuss open research challenges and opportunities.

INTRODUCTION

Sparked by recent advances in machine learning, lawmakers are reacting to the increasing dependence on automated decision-making with protective regulations, such as the General Data Protection Regulation of the European Union. These laws prescribe that decisions based on fully automated algorithms need to provide clear-cut reasonings and justifications for affected people. Hence, to address this demand, the field of E**x**plainable **A**rtificial **I**ntelligence (XAI) accelerated, combining expertise from different backgrounds in computer science and other related fields to tackle the challenges of providing logical and trustworthy decision reasonings.

The act of making something explainable entails a *process* that reveals and describes an underlying phenomenon. This has been the subject of study and research of different fields over the centuries. Therefore, to establish a solid foundation for explainable artificial intelligence, we need a structured approach based on insights, as well as well-studied practices on explanation processes. Structuring these methodologies and adapting them to the novel challenges facing AI research is inevitable for advancing effective XAI.

Position Statement: Multiple research domains established well-studied processes to communicate information and knowledge. Learning and transferring these processes will enable us to build effective methodological foundations for XAI.

Due to space constraints, this section compactly describes the main ideas gathered from our analysis of the related work and is based on **up to three of the most relevant research articles** for each of the six fields. A more complete overview of other works is provided in the appendix of this paper.

Methodology: Based on the analysis of the six research areas, we derived a number of different effective explanation strategies and best practices. These are grouped and categorized (over three iterations) into different elements, which are then put together to build the proposed model for explanation processes in XAI. In every step, the integrity of the model is cross-referenced with the original strategies (extracted from the research areas) to ensure their compatibility. The resulting conceptual model is described in the next section.

In this paper, we contribute a conceptual framework for effective explanation processes based on the analysis of strategies and best practices in six different research fields. **We postulate that accelerating the maturation of XAI and ensuring its effectiveness has to rely on the study of relevant research domains that established well-studied processes to communicate information and knowledge.** Consequently, learning and transferring these processes will bootstrap XAI and advance the development of tailored methodologies based on challenges unique to this young field.

Background

To place our model in context, we studied and analyzed related work from six different research areas: Pedagogy, Storytelling, Argumentation, Programming, Trust-Building, and Gamification.

Pedagogy. Proper methods of Pedagogy develop insight and understanding of how to do it [1]. However, good education involves many different strategies, such as induction and deduction [1, 2], methods, and mediums. Some methods, for instance, are explicit explanations using examples [1], group work and discussion [3], and self-explaining of students to students [3].

Storytelling in combination with data visualization is often practiced to explain complex phenomena in data and provide background knowledge [4]. Various strategies, e.g., martini glass structure, interactive slideshow, and drill-down story, exist to structure the narratives [5].

Dialog and Argumentation. In the humanities, many models for dialog and argumentation exist. Fan and Toni argue that "argumentation can be seen as the process of generating explanations" [6] and propose a theoretic approach [6]. Miller [7] provides an extensive survey of insights from the social sciences that can be transferred to XAI. Madumal et al. [8] propose a dialog model for explaining artificial intelligence.

Programming. Programming languages are inherently structural, and software should be self-explanatory. A popular and consent way to achieve this are design-patterns [9] and other concepts [10]. Design-patterns imply abstraction which is crucial for XAI.

Trust Building. There exists no active trust building scheme for AI. Instead, trust relies on explanation and transparency of the system [11]. Miller et al. and Siao et al. argue that the system has to continuously clear doubts over time to increase the user's trust, and that factors such as reliability and false alarm rate are essential for AI systems [12, 13].

Gamification is an integration of game elements and game thinking in non-gaming systems or activities with the goal to motivate the users and foster their engagement [14, p. 10]. To support the defined goal, the interaction of the system needs to be adapted "to a given user with game-like targeted communication" [15]. Such systems are usually designed to have several levels or modes with increasing complexity [14, p. 11]. It is important, though, that the user can specify which task to realize as next [16].

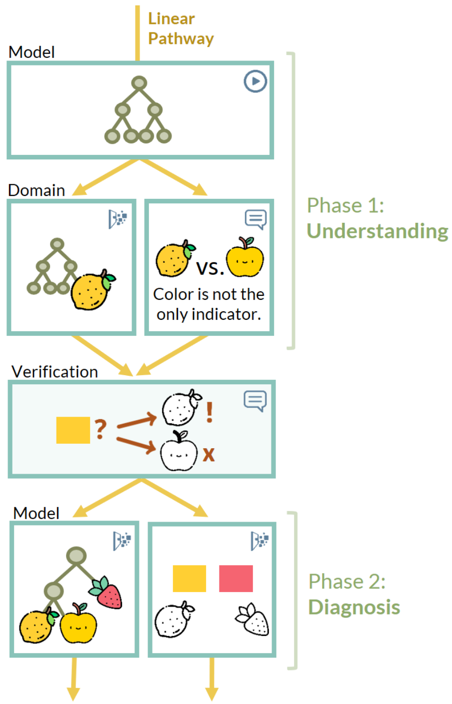

Figure 2: Fruit classification example: the first phase starts with a module that explains a decision tree classifier using a video. This transitions to two alternative explanation blocks where the left one uses visualizations of the model and the right block demonstrates the features using verbalization. The following verification block ensures that the previously learned material is understood. The second phase induces another two explanation blocks to diagnose the model. The left follows the visualization mantra of phase one. The right block uses a visualization to depict the data with its features.

MODELLING OF EXPLANATION PROCESSES IN XAI

We define an explanation process as a sequence of **phases** that, in turn, consist of **explanation** and **verification blocks**. Each of these **building blocks** uses a **medium** and a **strategy** for explanation or verification, respectively. The connection between these blocks is defined through **pathways**. A schematic representation of our model is depicted in Figure 1.

In this example, the explanation process is comprised of two phases (for AI understanding and diagnosis, respectively). The first phase consists of three explanation blocks, followed by a verification block. A linear pathway o→o→o connects all building blocks in the process. However, some explanation blocks are positioned on alternative paths. For example, users start the explanation in Phase 1 watching a video that uses a simplification strategy ✿⟶o for explanation; then they can choose between a visualization or a verbalization component as a second step. After choosing, for example, the verbal, explanation by abnormality o⟷o block, they can transition to a verification block, which uses a flipped classroom strategy ? ⟶ ! and verbalization as a medium.

Our *explanation process model* is instantiated in Figure 2 as a simplistic example of *explaining the classification of fruits* (using an analogous structure to the abstract process of Figure 1).

In the following, all elements of our model are discussed in more detail.

Pathways. An explanation process is comprised of different modules, i.e., phases that contain explanation and verification blocks. To connect these modules into a global construct, we define transitions, the so-called pathways. These can be **linear or iterative**, allowing building blocks in the process to be visited once or multiple times. Additionally, the navigation defined by them can be **guided or serendipitous**, enabling a strict framing or open exploration.

Mediums. Lipton [17] states that common approaches to describe how a model behaves and why, usually include verbal (natural language) explanations, visualizations, or explanations by example. In the current explainable AI systems, visualization is the most frequently applied medium. However, Sevastjanova et al. [18] argue for a combination of visualization and verbalization techniques, which can lead to deeper insights, and a better understanding of the ML model. For instance, the user could engage with an agent through a dialog system, by interacting with visualization and stating questions in natural language in order to understand the decisions made by the model. In storytelling, a combination of text and visual elements are used in diverse formats to communicate about the data effectively. Comics, illustrated texts, and infographics are three widely applied formats, which differ in the level of user guidance, and the way how text and visual elements are aligned [19]. In addition to the previously mentioned mediums, one might employ multimedia (e.g., video, audio, image, video game) to either engage the user on exploring the ML model in more detail (e.g., if the explanation is an integral part of a video game), or to provide explanations from another perspective.

Pathways

- **Linear** vs. **Iterative**
- **Guided** vs. **Serendipitous**

Mediums

- **Visualizations**
- **Verbalizations**
- **Infographics**
- **Illustrated Text**
- **Comics**
- **Videos**
- **Audios**
- **Images**
- **Video Games**
- **Dialog Systems**

Explanation Strategies

- **Inductive (Bottom Up)**
  - simplification
  - metaphorical narrative
  - divide and conquer
  - explanation by example
  - dynamic programming
  - depth first - breadth first
  - describe and define
  - teaching by categories
- **Deductive (Top Down)**
  - transfer learning
  - teaching by association
  - overview first, details on demand
  - drill down story
  - define and describe
- **Contrastive (Comparison)**
  - opposite and similar
  - example by abnormality

Verification Strategies ? ⟶ !

- flipped classroom
- reproduction
- transfer

Explanation Strategies. In logic and philosophy, often two opposing strategies for reasoning are named [2]. These are called *inductive* and *deductive* reasoning. The first strategy, **inductive reasoning**, is defined by Aristotle as "the conclusion process for a general knowledge out of observed events" [2]. The second strategy, **deductive reasoning**, builds the opposite and is defined as "the conclusion process from given premises to a logical closure" [2]. Such basic strategies can be found throughout literature in different fields, i.e., inductive (bottom-up) explanations vs. deductive (top-down) approaches. Inductive strategies first explain smaller and observable details, followed by complex relations. Hence, the explanation of the details should facilitate the understanding of the general and abstract concept. Examples of inductive strategies include; *simplifications, explanation by example*, or *metaphorical narratives*. Deductive strategies start with the whole picture (general idea) as an overview, then more details get added and explained to show a more complete view. Examples of deductive strategies include; *overview first, details on demand* or *transfer learning*.

In addition to these two groups, we identified another useful explanation method based on comparative analysis, so-called **contrastive explanation**. Such strategies rely on putting two phenomena side-by-side in a comparison and showing off their contrast. The explanation could then be performed using induction or deduction. One noteworthy example for this category is the strategy "*explaining by abnormality*" where the unusual manifestations of a phenomenon is shown to contrast the "normal" state and prevent misconceptions.

Lastly, it is with noting that the overall structure of the phases in an explanation process can be designed based on guidelines derived from explanation strategies (optionally increasing the complexity of the process to more intricate or recursive explanations).

Verification Strategies. To ensure that users have gained an encompassing and sound understanding of the underlying subject matter, explanation processes need to include verification strategies. We propose optional verification blocks at the end of each phase to establish a stable common ground as a conclusion for that phase, before allowing users to advance to the next one (typically increasing the complexity). In contrast to explanation strategies, verification strategies usually require users to demonstrate the learned phenomena. They include strategies that are based on questions for *reproducing* or *transferring* knowledge, as well as "*flipping the classroom*", i.e., having users explain to the system the learned concepts.

DISCUSSION: BEST PRACTICES, GUIDELINES, AND RESEARCH OPPORTUNITIES

Several considerations have to be made to select and structure the presented strategies. The decisions should be mainly based on (i) the targeted level of detail; (ii) the target audience; (iii) the desired level of interactivity of the target audience. The level of detail considerably impacts the choice of the strategies and their respective structure and sequence. The spectrum ranges from answering the question of what the respective machine learning model(s) are achieving to how the model(s) work in detail. To answer the former, a possible consideration could be the use of metaphorical narratives [20] while the latter needs more accurate and precise descriptions conveyed through mathematical notations and pseudocode. Two things have to be considered regarding the target audience. Their size and composition affect the use of mediums and the level of interactivity while their background knowledge on the subject matter reduces the required distance to reach the desired level of detail. The interactivity may vary during the phases. It is beneficial to increase the level of interactivity for verification phases to receive more and more profound feedback.

It is possible to engage users using gamification to raise their motivational support [14, p. 10] while continuously receiving and providing feedback to the explanatory and the user [15]. The feedback aspect can be well exploited in verification blocks whereas the motivational support may drive the user to explore multiple *pathways* of the explanation process as well as exploring the machine learning model itself in more detail. Tracking and displaying the progress serves as an extrinsic motivation [14, p. 52] allowing the user to better navigate the various *pathways*.

CONCLUSIONS

Valuable strategies can be extracted and abstracted from varying research areas. These strategies serve an important and well-researched baseline to bootstrap the process of explainable machine learning. Our proposed model classifies these strategies and combines them as building blocks to actualize an explanation process for machine learning while keeping the flexibility of using different mediums and transitioning paths. The list of collated strategies is not inclusive, yet the proposed model allows many variations and extensions which provides space for further research opportunities. Additionally, existing XAI approaches can be analyzed and deconstructed to extract the building blocks to validate whether our proposed model can be adopted. Successful explanation processes can then be compared and analyzed regarding common patterns.

Research Opportunities

- Implement, test, and verify different explanation strategies. Is the knowledge from other domains transferable to XAI explanation processes?
- Identify the most suitable processes for different settings, tasks, and AI models.
- Extend strategies to tailor them to XAI.
- Evaluate the proposed explanation model, e.g., through user studies and texting out of alternative models.
- Make XAI processes reactive to the users' interaction through automatic pathway generation, e.g., through active learning.
- Tailoring the explanation strategies: which strategy works best in which environment and for specific target groups, as well as various levels of AI complexity.
- Designing Visual Analytics systems that integrate the users' interactions into a mixed-initiative model-refinement cycle.

Take Home Messages

- Studying and combining explanation processes is critical to establish effective XAI methodologies and mature this research field.
- Best practices and tailored explanation processes can streamline XAI and account for different circumstances, such as task complexity, data characteristics, model type, and user expertise.
- Given clear problem specifications, as well as well-studied and detailed guidelines, we can progress toward automatically generating XAI processes as design templates for successful explanations and model refinements.

REFERENCES

[1] Odora Ronald James. 2014. Using Explanation as a Teaching Method: How Prepared are High School Technology Teachers in Free State Province, South Africa. In *Journal of Social Sciences*, 38:71–81.
[2] Robert J. Sternberg and Karin Sternberg. 2016. *Cognitive psychology*. Nelson Education.
[3] Richard Gunstone, editor. 2015. *Explaining as a Teaching Strategy. Encyclopedia of Science Education*. Springer Netherlands, Dordrecht, 423–425.
[4] Robert Kosara and Jock D. Mackinlay. 2013. Storytelling: The Next Step for Visualization. *IEEE Computer*, 46, 5, 44–50.
[5] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. Vis. Comput. Graph.*, 16, 6, 1139–1148.

[6] Xiuyi Fan and Francesca Toni. 2014. On Computing Explanations in Abstract Argumentation. In *Proc. of the ECAI*. IOS Press, 1005–1006.

[7] Tim Miller. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artifical Intelligence*, 267, 1–38.

[8] Prashan Madumal, Tim Miller, Frank Vetere, and Liz Sonenberg. 2018. Towards a Grounded Dialog Model for Explainable Artificial Intelligence. In *Workshop on SCS*.

[9] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. 1995. *Design Patterns: Elements of Reusable Object-oriented Software*. Addison-Wesley Longman Publishing.

[10] Mark Dominus. 2006. Design patterns of 1972 — The Universe of Discourse. [Online; accessed 7-February-2019]. (2006). https://blog.plover.com/prog/design-patterns.html.

[11] Wolter Pieters. 2011. Explanation and Trust: What to Tell the User in Security and AI? *Ethics and information technology*, 13, 1, 53–64.

[12] Keng Siau and Weiyu Wang. 2018. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, 31, 47–53.

[13] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of Inmates Running the Asylum. *CoRR*, abs/1712.00547.

[14] Karl M. Kapp. 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. (1st edition). Pfeiffer & Company.

[15] Cathie Marache-Francisco and Eric Brangier. 2013. Process of Gamification. *Proceedings of the 6th Centric*, 126–131.

[16] Jakub Swacha and Karolina Muszynska. 2016. Design Patterns for Gamification of Work. In *Proc. of TEEM*.

[17] Zachary Chase Lipton. 2016. The Mythos of Model Interpretability. *CoRR*, abs/1606.03490.

[18] Rita Sevastjanova, Fabian Beck, Basil Ell, Cagatay Turkay, Rafael Henkin, Miriam Butt, Daniel A. Keim, and Mennatallah El-Assady. 2018. Going beyond Visualization: Verbalization as Complementary Medium to Explain Machine Learning Models. In *Proc. of VISxAI Workshop, IEEE VIS*.

[19] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proc. of ACM CHI*.

[20] Wolfgang Jentner, Rita Sevastjanova, Florian Stoffel, Daniel A. Keim, Jürgen Bernard, and Mennatallah El-Assady. 2018. Minions, Sheep, and Fruits: Metaphorical Narratives to Explain Artificial Intelligence and Build Trust. In *Proc. of VISxAI Workshop, IEEE VIS*.

APPENDIX: RELATED WORK

Complete overview of surveyed fields, setting the most relevant related work in context.

Pedagogy

"Good teaching is good explanation." [21] Proper methods of Pedagogy develop insight and understanding of how to do it [1]. However, good education involves many different strategies, like induction and deduction [1, 2], methods, and mediums. Some methods, for instance, are explicit explanations using examples [1], group work and discussion [3], and self-explaining of students to students [3]. These methods and strategies include the logic and philosophical base with induction and deduction [1, 2]. Further logical operations can be incorporated to extend these methods and strategies, namely comparison, analysis, synthesis, and analogy. [1] In this context, three different parts of an explanation exist, something that is to be explained, an explainer, the one who explains, and the explainee, who gets the explanation. [22] If the explainer wants to provide a good explanation to the explainee, the explanation has to be clearly structured and interesting to the explainer [23]. Good explanations can invoke understanding. However, bad explanations may lead to confused and bored explainee and explainer [23]. Brown and Atkins [24] describe three types of explanations: descriptive, interpretive, and reason giving. A descriptive explanation can be defined as describe and define and explains the processes and procedures [24]. An interpretive explanation specifies the central meaning of a term and can be seen as define and describe. And last, reason giving explanation shows reasons based on generalizations and can be interpreted as teaching by categories. There are more proposed strategies and methods, e.g. Wragg [23] or Brown and Armstrong [25], but in general, they can be summarized with the explanation strategies and methods above.

Storytelling

Storytelling has been used for millennia in human history to communicate information, transfer knowledge, and entertainment [5]. Outlining the complete field is almost impossible as storytelling is as diverse as humanity. However, commonalities appear when looking at this field at a more abstract level. We explicitly focus on works of storytelling in combination with data visualization as this is often practiced to explain complex phenomena in data and provide background [4]. Machine learning follows this goal, however, it is not sufficient to understand the phenomena in the data. A user must also learn about the reason how and why this phenomena appears to verify and validate the phenomena. This further affects the trust building process positively. Various strategies exist to structure narratives uniting data visualization [5] and best practices have been extracted and summarized to improve storytelling for visualizations [26]. We transfer and provide a taxonomy for these strategies that we deem useful to explain machine learning.

Dialog and Argumentation

In the humanities, many models for dialog and argumentation exist, with Hilton stating that "causal explanation takes the form of conversation" [27]. This conversation involves both cognitive and social processes [8] and is "first and foremost a form of social interaction" [7]. According to Grice, utterances in this conversation should follow the four maxims of *quantity*, *quality*, *relation* (relevance) and *manner* [28]. Miller notes that "questions could be asked by interacting with a visual object, and answers could similarly be provided in a visual way." [7] Many of these principles have been applied to explainable AI, as surveyed by Miller [7]. Fan and Toni argue that "argumentation can be seen as the process of generating explanations" [6] and propose two theoretic approaches [6, 29]. Madumal et al. propose a dialog model for explaining artificial intelligence [8], and Zeng et al. introduced an argumentation-based approach for context-based explainable decisions [30]. Here, the "schemes for practical reasoning" and "schemes for applying rules to cases" from Walton and Macagno's classification system for argumentation schemes [31] seem particularly interesting.

Programming

Typically, during programming, common software design patterns and best-practices are followed. While the main goal of such patterns is to provide "general, reusable solution[s] to [. . .] commonly occurring problem[s]" [32], they often act as a self-explanation strategy for complex software systems. For the programmer, software design patterns improve readability, traceability and help by building up a mental model of the system. Many software design patterns can be classified using the categories introduced in section **Modelling of Explanation Processes in XAI**. The program flow is the *pathway* of code: it can be *linear* (block), *iterative* (loop) or *part of itself* (recursion). Algorithms can follow a *top-down* or *bottom-up* approach. The main strategy followed in software design patterns is *abstraction*. Abstraction is the core concept of modern high-level programming languages [33] and is closely related to Shneiderman's mantra "overview first, [. . .] details on demand" [34]. While, again, the strategy in the first instance has a practical use, it also takes an explanatory role for the programmer. He can understand the full program on a higher level and, if needed, can go deeper to view the details. Abstraction does not only occur as a concept of language design, but also in many discrete programming patterns. This ranges from the simple concept of subroutines [10] up to many of the design patterns for object-oriented programming proposed by Gamma et al. (GoF) [9], e.g. *facade* or *iterator*.

Trust Building

Trust in machine learning systems is highly dependent on the system itself and how it can be explained. Glass, Mcguinness, and Wolverton find that "trust depends on the granularity of explanations and transparency of the system" [11, 35]. As our trust building is usually done with humans, many natural trust mechanism rely on a person's presence. However, these are not available in AI systems, and therefore the explanation and transparency of the system become the most important factors that influence the user's trust. Pieters argues there is a difference between having trust a system, where the user completely understands the decisions of the system itself and therefore make active decisions about the result, and confidence in a system, where the user does not need to know inner workings in order to use its results [11]. Miller, Howe, and Sonenberg and Siau and Wang argue that trust is dynamic and build up in a gradual manner. Furthermore there is a differentiation between initial trust that the user has obtained through external factors, e.g. cultural aspects, and the trust that he builds while using the system. The system should continuously clear potential doubts over time by providing additional user-driven information. Factors such as reliability, validity, robustness, and false alarm rate influence how the user develops trust in the system, and should play an integral role when designing the system [12]. Lombrozo shows that the people disproportionately prefer simpler explanations over more likely explanations. Therefore explanations should aim to only carry the appropriate amount of information. Furthermore, they found that people prefer contrastive explanations at certain parts of the system, because otherwise the cognitive burden of a complete explanations is too great [13].

Gamification

Gamification is an integration of game elements and game thinking in non-gaming systems or activities [14, p. 10]. It aims at motivating users [37, p. 4] to foster their engagement [14, p. 10]. Gamification uses several concepts to achieve this goal. Usually, a user is asked to accomplish tasks to earn points; these points are accumulated and based on the achieved result the user may receive rewards. To support the defined goal, the interaction needs to be adapted "to a given user with game-like targeted communication" [15]. Thus, to increase engagement, games are usually designed to have several levels or modes with increasing complexity [14, p. 10]. It is important, though, that the user can specify which task to realize as next and which *pathway* to take to achieve the goal [16]. In order to make these systems more attractive, game elements are designed to generate positive emotions. Usually, it is done by applying a specific vocabulary (e.g., *simplification*) or narrations [15]. According to Bowser et al. [38], different user groups prefer a different type of interface. Important is, thus, to adapt the system to the specific user profile, by showing only the elements relevant for his particular task [15].

REFERENCES

[1] Odora Ronald James. 2014. Using Explanation as a Teaching Method: How Prepared are High School Technology Teachers in Free State Province, South Africa. In *Journal of Social Sciences*, 38:71–81.

[2] Robert J. Sternberg and Karin Sternberg. 2016. *Cognitive psychology*. Nelson Education.

[3] Richard Gunstone, editor. 2015. *Explaining as a Teaching Strategy*. *Encyclopedia of Science Education*. Springer Netherlands, Dordrecht, 423–425.

[4] Robert Kosara and Jock D. Mackinlay. 2013. Storytelling: The Next Step for Visualization. *IEEE Computer*, 46, 5, 44–50.

[5] Edward Segel and Jeffrey Heer. 2010. Narrative Visualization: Telling Stories with Data. *IEEE Trans. Vis. Comput. Graph.*, 16, 6, 1139–1148.

[6] Xiuyi Fan and Francesca Toni. 2014. On Computing Explanations in Abstract Argumentation. In *Proc. of the ECAI*. IOS Press, 1005–1006.

[7] Tim Miller. 2019. Explanation in Artificial Intelligence: Insights from the Social Sciences. *Artifical Intelligence*, 267, 1–38.

[8] Prashan Madumal, Tim Miller, Frank Vetere, and Liz Sonenberg. 2018. Towards a Grounded Dialog Model for Explainable Artificial Intelligence. In *Workshop on SCS*.

[9] Erich Gamma, Richard Helm, Ralph Johnson, and John Vlissides. 1995. *Design Patterns: Elements of Reusable Object-oriented Software*. Addison-Wesley Longman Publishing.

[10] Mark Dominus. 2006. Design patterns of 1972 — The Universe of Discourse. [Online; accessed 7-February-2019]. (2006). https://blog.plover.com/prog/design-patterns.html.

[11] Wolter Pieters. 2011. Explanation and Trust: What to Tell the User in Security and AI? *Ethics and information technology*, 13, 1, 53–64.

[12] Keng Siau and Weiyu Wang. 2018. Building Trust in Artificial Intelligence, Machine Learning, and Robotics. *Cutter Business Technology Journal*, 31, 47–53.

[13] Tim Miller, Piers Howe, and Liz Sonenberg. 2017. Explainable AI: Beware of Inmates Running the Asylum. *CoRR*, abs/1712.00547.

[14] Karl M. Kapp. 2012. *The Gamification of Learning and Instruction: Game-based Methods and Strategies for Training and Education*. (1st edition). Pfeiffer & Company.

[15] Cathie Marache-Francisco and Eric Brangier. 2013. Process of Gamification. *Proceedings of the 6th Centric*, 126–131.

[16] Jakub Swacha and Karolina Muszynska. 2016. Design Patterns for Gamification of Work. In *Proc. of TEEM*.

[17] Zachary Chase Lipton. 2016. The Mythos of Model Interpretability. *CoRR*, abs/1606.03490.

[18] Rita Sevastjanova, Fabian Beck, Basil Ell, Cagatay Turkay, Rafael Henkin, Miriam Butt, Daniel A. Keim, and Mennatallah El-Assady. 2018. Going beyond Visualization: Verbalization as Complementary Medium to Explain Machine Learning Models. In *Proc. of VISxAI Workshop, IEEE VIS*.

[19] Zezhong Wang, Shunming Wang, Matteo Farinella, Dave Murray-Rust, Nathalie Henry Riche, and Benjamin Bach. 2019. Comparing Effectiveness and Engagement of Data Comics and Infographics. In *Proc. of ACM CHI*.

[20] Wolfgang Jentner, Rita Sevastjanova, Florian Stoffel, Daniel A. Keim, Jürgen Bernard, and Mennatallah El-Assady. 2018. Minions, Sheep, and Fruits: Metaphorical Narratives to Explain Artificial Intelligence and Build Trust. In *Proc. of VISxAI Workshop, IEEE VIS*.

[21] Robert C. Calfee. 1986. *Handbook of Research on Teaching*. Macmillan.

[22] Fairhurst MA. 1981. Satisfactory explanations in the primary school. *Journal of Philosophy of Education*, 15, 2, 205–213.

[23] Wragg EC and Brown G. 1993. *Explaining*. Routledge Publishers.

[24] 1997. *Explaining. The Handbook of Communication Skills*. Routlege Publishers, 199–229.

[25] 1984. *Explaining and explanations. Classroom Teaching Skills*. Nichols Publishing Company, 121–148.

[26] Nahum D. Gershon and Ward Page. 2001. What Storytelling can do for Information Visualization. *Commun. ACM*, 44, 8, 31–37.

[27] Denis J. Hilton. 1990. Conversational processes and causal explanation. *Psychological Bulletin*, 107, 1, 65–81. DOI: 10.1037/0033-2909.107.1.65.

[28] Herbert Paul Grice. 1967. Logic and Conversation. In *Studies in the Way of Words*. Paul Grice, editor. Harvard University Press, 41–58.

[29] Xiuyi Fan and Francesca Toni. 2015. On explanations for non-acceptable arguments. In *Theory and Applications of Formal Argumentation*. Elizabeth Black, Sanjay Modgil, and Nir Oren, editors. Springer International Publishing, Cham, 112–127. ISBN: 978-3-319-28460-6.

[30] Zhiwei Zeng, Xiuyi Fan, Chunyan Miao, Cyril Leung, Chin Jing Jih, and Ong Yew Soon. 2018. Context-based and Explainable Decision Making with Argumentation. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (AAMAS '18). International Foundation for Autonomous Agents and Multiagent Systems, Stockholm, Sweden, 1114–1122. http://dl.acm.org/citation.cfm?id=3237383.3237862.

[31] Douglas Walton and Fabrizio Macagno. 2015. A Classification System for Argumentation Schemes. *Argument and Computation*, 6, 3, 219–245.

[32] Wikipedia contributors. 2019. Software design pattern — Wikipedia, the free encyclopedia. [Online; accessed 7-February-2019]. (2019). https://en.wikipedia.org/w/index.php?title=Software_design_pattern&oldid=879797369.

[33] Wikipedia contributors. 2019. High-level Programming Language — Wikipedia, The Free Encyclopedia. [Online; accessed 7-February-2019]. (2019). https://en.wikipedia.org/w/index.php?title=High-level_programming_language&oldid=879754477.

[34] Ben Shneiderman. 1996. The Eyes have it: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings 1996 IEEE Symposium on Visual Languages*. IEEE, Boulder, CO, USA, USA, (September 1996), 336–343. DOI: 10.1109/VL.1996.545307.

[35] Alyssa Glass, Deborah Mcguinness, and Michael Wolverton. 2008. Toward Establishing Trust in Adaptive Agents. In (January 2008), 227–236. DOI: 10.1145/1378773.1378804.

[36] Tania Lombrozo. 2007. Simplicity and probability in causal explanation. *Cognitive Psychology*, 55, 232–257.

[37] Yu-kai Chou. 2015. *Actionable Gamification: Beyond Points, Badges, and Leaderboards*. Octalysis Group Fremont, CA.

[38] Anne Bowser, Derek Hansen, and Jennifer Preece. 2013. Gamifying Citizen Science: Lessons and Future Directions. In *Workshop on Designing Gamification: Creating Gameful and Playful Experiences*.