Erschienen in: Workshop on Visualization for AI Explainability at IEEE

Minions, Sheep, and Fruits: Metaphorical Narratives to Explain Artificial Intelligence and Build Trust

Wolfgang Jentner[1], Rita Sevastjanova[1], Florian Stoffel[1], Daniel Keim[1], Jürgen Bernard[2], Mennatallah El-Assady[1,3]

[1]University of Konstanz, Germany
[2]TU Darmstadt, Germany
[3]University of Ontario Institute of Technology, Canada

Fig. 1. A transitive mapping of terms via a metaphorical narrative forces both parties, domain experts and modeling experts, to reduce the complexity of their vocabulary in order to map their individual mental models and domain understanding.

Abstract— Advanced artificial intelligence models are used to solve complex real-world problems across different domains. While bringing along the expertise for their specific domain problems, users from these various application fields often do not readily understand the underlying artificial intelligence models. The resulting opacity implicates a low level of trust of the domain expert, leading to an ineffective and hesitant usage of the models. We postulate that it is necessary to educate the domain experts to prevent such situations. Therefore, we propose the *metaphorical narrative* methodology to transitively conflate the mental models of the involved modeling and domain experts. Metaphorical narratives establish an uncontaminated, unambiguous vocabulary that simplifies and abstracts the complex models to explain their main concepts. Elevating the domain experts in their methodological understanding results in trust building and an adequate usage of the models. To foster the methodological understanding, we follow the Visual Analytics paradigm that is known to provide an effective interface for the human and the machine. We ground our proposed methodology on different application fields and theories, detail four successfully applied metaphorical narratives, and discuss important aspects, properties, and pitfalls.

1 Introduction

Artificial Intelligence (AI) successfully solves many complex problems and significantly impacts our everyday life. One primary goal of many researchers is to make AI accessible to broader user groups, not only to modeling experts but also to domain experts and practitioners in various data-centered application domains. However, many AI models are very complex and difficult to understand, even for modeling experts. Example classes for complex machine learning and AI models include clustering [13], dimensionality reduction [25], regression models [22], or classification [21] (particularly deep neural networks [18] and related techniques). Explaining such complex AI models is a challenging task, especially if the targeted user group involves domain experts. It is comprehensible that many domain experts still have concerns when an unknown AI model is to be adopted into their working practices or data-centered workflows. In line with related works in several domains [3, 12, 29], we observed that domain experts partially refused to adopt new paradigms in data science (enabled with AI) but still perform at least parts of their workflow with general purpose tools for data analysis like Tableau or Excel. We assume that this missing *trust* in complex AI models is the result of either a lack of *model understanding* or the ability of *model validation*, or both.

To overcome problems related to missing trust-building, new methodologies are needed to better explain complex AI models, and enable domain users to ease the access and interaction with AI models. It is the obligation of modeling experts to close the gap between the complexity of AI models on the one hand and the abilities of domain experts to gain a model understanding on the other hand. This mediating task [1] is crucial in the design phase as the modeling expert must apply the domain expert's data, analysis tasks, and requirements in AI models to adequately support the domain experts in their data-centered endeav-

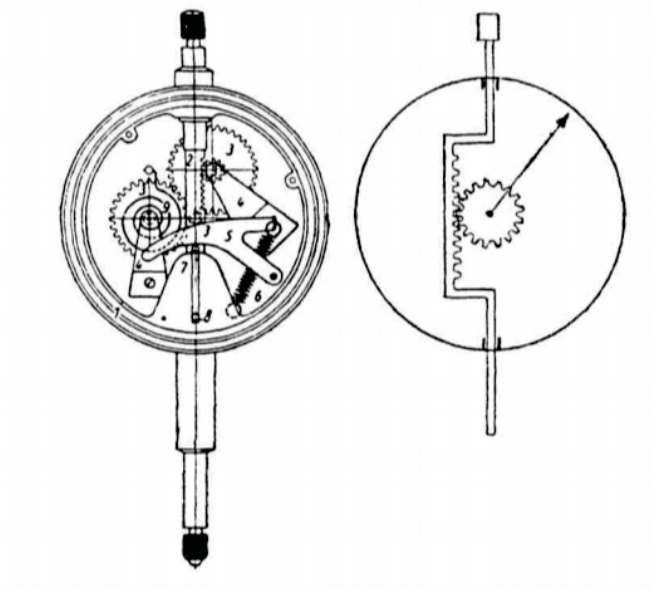

Fig. 2. An applied vertical didactic reduction. The left shows the actual mechanical parts of a dial indicator. Right: the complexity is reduced to explain the main functionality [9].

ors. After deployment, domain experts and also other users that were not involved in the design process must build trust in the system to understand its capabilities and limitations [19]. Transparency of the used models is key, and one crucial aspect is their explainability. However, a common problem in bridging the mental worlds of the domain experts and the modeling expert is known as the *curse of knowledge* [10]. This cognitive bias describes the problem that it is not trivial for experts understanding what knowledge the other party has and especially not has. Therefore, experts have difficulties sharing their knowledge with others.

In this work, we postulate that modeling experts indeed have possibilities to make AI more explainable. We contribute the methodology of *metaphorical narratives* to explain AI. We base this methodology and build upon best practices and theories from pedagogy, design study methodology, translation theory, and software interfaces. Metaphorical narratives establish an *uncontaminated vocabulary* that is used by two communication parties to conflate their individual mental models and domain understanding. This lowers the cognitive loads of both parties, enforces the abstraction and simplification of the concepts, and positively impacts the trust building process of the user. Trust-building can be achieved in two principal ways: by enhancing *model understanding* and through *model validation*. We propose metaphorical narratives as a complementary methodology to other well established approaches such as collaborative model building or simplified modeling [8].

In the next Section, we introduce best practices, principles, and theories from varying fields that serve as the basis for our method. Section 3 explains how the metaphorical narratives complement the Visual Analytics Process and describe their impact on the trust building process. Furthermore, we provide four exemplary cases where metaphorical narratives have been successfully applied to facilitate explainable AI for various types of AI models. Our methodology is discussed in Section 4. Section 5 draws the conclusions and postulates further research opportunities.

2 Towards Metaphorical Narratives

Our methodology builds upon best practices in several complementary application fields and theories. In particular, we build our solution towards explainable AI on didactic reduction, software interfaces, design study methodology, and translation theory.

Didactic Reduction The term "didactic reduction" was first established by Grüner [9] and is widely known in pedagogical theory. In English this is loosely translated by "simplification" or "elementarization" [7]. The didactic reduction is distinguished into the horizontal didactic reduction and the vertical didactic reduction. The first is a presentational reduction and describes the use of examples, analogies, and metaphors to produce a concrete illustration of the problem and the method that tries to resolve the problem. Examples for a horizontal reduction are the use of concrete values in a given formula and calculating the results or the use of the visual metaphor of a river network to describe the blood vessel system. The vertical didactic reduction means the reduction of content to simplify complex contexts by leaving out details that are irrelevant for the target audience. Figure 2 shows an illustrated vertical didactic reduction of a mechanical dial indicator. The left illustration depicts the actual mechanical parts of the dial indicator whereas the right picture reduces this complexity to explain the main functionality.

The main principles for didactic reduction are: (i) the technical correctness must be maintained meaning that the generalized and simplified concepts and laws remain valid; (ii) the reduced content can be extended to explain more specific concepts and more details can be added and (iii) the appropriateness of the content must be adjusted to the respective target audience (i.e., the domain expert(s)). Metaphorical narratives should obey these principles of didactic reduction.

Interfaces in Object-Oriented Software When following the object-oriented programming paradigm, interfaces are a well-known and established technique to abstract details of an implementation. Meyer introduces interfaces as a technique to realize one of the five rules of modularity, which is *information hiding* [20]. A motivation for information hiding is to ensure a separation of function from the actual implementation, which is in line with our idea of explaining the general function of an AI without providing implementation details. Inherently, this approach requires a level of abstraction to provide an interface to some external component. A positive side-effect of interfaces is *modularization* [30]. Metaphorical narratives implement the concept of information hiding through the abstraction and simplification process. A well-chosen metaphorical narrative can additionally contribute to the modularization effect which is especially important in the design phase as underlying models can be modified or even exchanged where the concepts explained through the metaphorical narrative remain valid.

Design Studies Research on user-centered design and design study methodology has considerably inspired the working practices in our field. In the formative phase of the design study, effectively characterizing and abstracting the domain problem is a key skill of design study researchers. This step mainly includes the mapping of problems and data from the vocabulary of the specific domain into a more abstract and generic description [23]. Learning the domain language seems to be a recommendation for successful visualization collaboration [16]. Sedlmair et al. emphasize the importance of learning just enough about a domain to abstract rather than understanding all details [27]. In contrast to gathering a full-blown understanding of the target domain, it is more important to find a sweet spot about how much domain knowledge to acquire before becoming ineffective. Taking the perspective of domain experts, van Wijk argues that domain experts can not even be expected to know the concepts of design study researchers: in an ideal world, they would not even need to know [31].

One strategy to meet a sweet spot between domain familiarization and abstraction on both directions of the collaboration is the commitment to a common vocabulary [28], or to risk the drift [11] and surpass the own disciplinary boundaries to reach the mental space of the other field. Metaphorical narratives are one method to establish a simplified, common vocabulary and play an important role in different phases of the design study process. Especially in the formative phase [27] including various design iterations and validations, metaphorical narratives may define the required behavior of a model that is to be designed (narrative $\rightarrow$ model), or be used to describe the resulting functionality of a model that was designed (model $\rightarrow$ narrative). Reflecting our experiences of previous design studies, metaphorical narratives themselves also develop and improve in the course of the design process. As such,

metaphorical narratives can even be seen as an additional end product from design study projects.

Translation Theory Visualization researchers and their domain experts arguably speak different languages. It cannot be guaranteed that the two collaborating groups understand all important aspects of their respective counterpart. Hence, concepts, ideas, and explanations often get "lost in translation." This metaphor of different domains expressing their mental models in different ways builds a communication barrier that can be directly mapped and understood by *translation theory* [4]. Translation theory describes the phenomena that occur when mapping concepts from one language to another. For natural languages, these phenomena include an imprecise mapping between languages, by definitions, as each language has its unique characteristics, structures, and conceptual abstractions. For example, one language might have a differentiation between two concepts that the other does not have (concept or term ambiguity); one word ("gift") might be written in the same way but invoke different meanings for different audiences (English: present, German: poison); or a language might have a notion (German: "Ohrwurm" – literary, *ear-worm* - meaning: "song stuck in the head") that does not have an adequate equivalent in the other language. All of these phenomena are known to linguists and can lead to miscommunication. However, these obstacles are also observable in between communications of different disciplines and can hamper the understanding of abstract and complex processes in visual analytics.

A straightforward method to overcome such issues is for one of the two parties to learn the concepts and mental models of the other party in order to map and express complex thoughts in the language spoken by the counterpart. In practice, however, this might lead to even more miscommunication due to the fact that the person going beyond their comfort zone might not use the correct terminology when translating ideas. In addition, learning the language of another domain is a tedious process that is often not feasible for conducting collaborations efficiently [27]. Hence, while out of their depth, one partitioner group might think that the usage of two particular terms is interchangeable and introduce wrong associations during the explanation.

Metaphorical narratives can be used as a middle ground for mapping concepts from the two communication parties and establish a neutral "third language". Hence, avoiding the pitfall of letting one party bare the burden of translating their complex mental models to the other language, we propose to share this responsibility and let both parties think around the corner. In addition, using such an approach makes the process of mapping and simplifying the concepts to such narratives an intentional and active process. Hence, the liability of ensuring the validity of this mapping becomes a shared endeavor that both parties have to guarantee for their respective side.

3 Metaphorical Narratives

Metaphorical narratives establish a simple vocabulary that is distinct to the vocabularies of the domain expert and the modeling expert. All terms that are being used in the metaphorical narrative are unambiguous in their semantics. The number of terms, as well as their complexity, is reduced in comparison to the domain-specific concepts and is just high enough to explain the main functionality of the concepts. This requires that the mapping from the specific domains to the metaphorical narrative is indeed ambiguous (Figure 3). Metaphorical narratives combine the insights, principles, and best practices that we can observe in the aforementioned domains, i.e., from didactic reduction, software interface design, design study methodology, and translation theory. In the following, we associate the methodology of metaphorical narratives to the the visual analytics process as well as to the trust-building process of domain experts. Finally, we describe four exemplary cases where metaphorical narratives have successfully been developed and applied in collaborative projects to foster explainable AI.

3.1 Visual Analytics

Visual analytics combines the strength of AI and visualization. The visual analytics process [15, 24] builds a valuable basis for explainable AI by using metaphorical narratives. The illustration in Figure 3 combines the visual analytics process with our methodology of metaphorical narratives. We emphasize the fact that modeling experts and domain experts have different backgrounds when looking at a view showing visualizations of AI (thought bubbles). Domain experts understand the domain concepts that are encoded in the data, as well as the tasks and problems existing in the respective domain. Modeling experts have a deep understanding of the concepts as they select, combine, and tune the underlying used AI models as well as visualization techniques. Metaphorical narratives establish a mediating vocabulary such that the concepts of either party can be transitively mapped. The so established common ground serves as a basis to communicate the essential functional behavior of the AI models and visualizations. The domain experts, on the other hand, can use the metaphorical narrative to explain the data, artifacts in the data that may lead to some effect in the output of the system, and also the desired tasks that they try to accomplish. The domain experts in their role as the users of the system benefit from the metaphorical narrative as it fosters the familiarization with the system, especially in the beginning. Along these lines, the narratives also prevent either of the practitioners falling into domain-specific jargons, terminologies, and practices. Finally, the metaphorical narrative can be used by both parties for analytical reasoning and sensemaking.

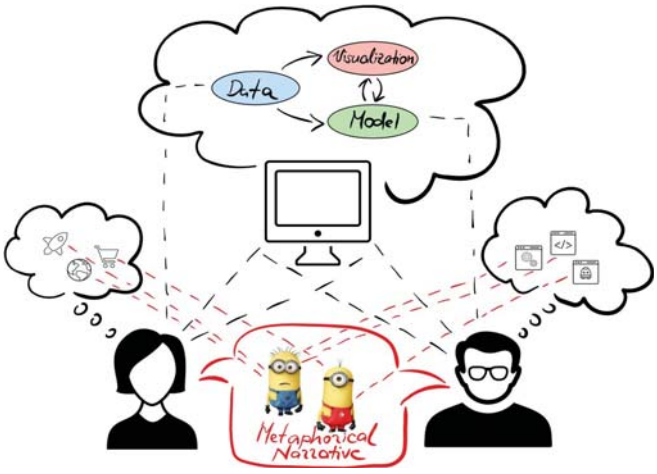

Fig. 3. Conflation of the Visual Analytics process with our methodology of metaphorical narratives. Modeling experts and data experts have different backgrounds, but benefit from a common visual vocabulary, making AI more explainable, e.g., in collaborative endeavors.

All of these are important in a visual analytics' setting as the user's involvement in the steering and decision making during the analysis process is desired. The user must, therefore, be able to understand the basic underlying processes to effectively exploit the system. On the other hand, visual analytics is well suited to support metaphorical narratives as it can convey the metaphorical narratives through the visualizations and graphical user interfaces.

3.2 Trust-Building Model

The application of metaphorical narratives positively influences the trust-building process of the domain expert. We argue that the trust-building of a domain expert in an AI model can be decomposed into two major dimensions. We hereby assume that the user has a general motivation to work with the application as it promises to ease her daily routines and provide more insights into some available data.

Our proposed trust-building model is depicted in Figure 4. The dimension, shown on the x-axis, is called *expectation match*. Typically, domain experts have a good understanding of what to expect as the outcome of some given system according to their expertise. We denote the *expectation match* as two intersecting sets whereas set M represents the output of the system and D the output as it is expected by the

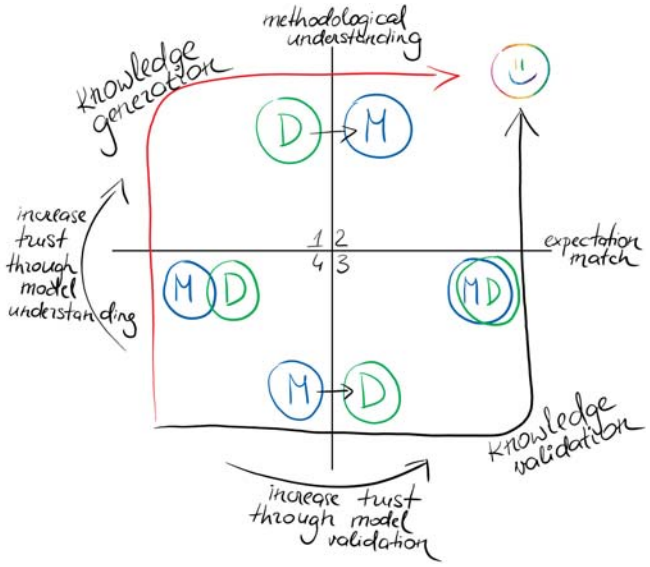

Fig. 4. Our proposed trust-building model illustrates the trust-building process using two orthogonal concepts: *methodological understanding* and *expectation match*.

domain expert. The expectations arise from the respective domain knowledge of the task and the data that are provided by the domain expert. An increasing *expectation match* is visualized in the chart from left to right. Quantifying this dimension is not trivial due to the facts that: (i) it is often difficult for domain experts to fully formalize their expectations and (ii) the output of any system is typically not consumed directly but through interpreting different (interactive) visualizations whereas the interpretation is affected by many occurring biases [6]. An advancement to the right of the chart can be performed in two ways. The first one is to modify the model such that the output of the model changes. We denote this as $M \rightarrow D$. The second way is an adaption of the user's expectations which we refer to as $D \rightarrow M$. Both are not exclusive and may happen simultaneously in practice.

The dimension depicted on the y-axis represents the *methodological understanding*. We hereby explicitly refer to the complete system including all used AI models plus the visualizations and interaction possibilities. Furthermore, the *methodological understanding* refers to an understanding of what the system as a whole is performing, how the data is transformed during the process, and how this is associated with the given task(s). A position at the bottom thus depicts a user with no *methodological understanding* something which is also typically referred to as a black box. The other extreme in the upper region of the chart is a user that has a full understanding of how the data is transformed and how the results are being generated and can be explained.

The resulting four quadrants describe states of the user concerning the system and can explain how the user possibly reacts. Quadrant 4 describes an user with no *methodological understanding* and no *expectation match*. Through interviews and observations of the domain experts, a typical reaction in such a case is the repetition of the analysis process to see whether the output of the system is changing or not. This may also include various, random parameter settings. However, if the output does not increase the *expectation match*, the users discontinue using the system (or the respective part of the system) and explore alternative ways to receive the expected output. This might be even to an extent where the data is processed manually. We, therefore, consider this a state where a user has no trust in the system.

Quadrant 3 refers to a domain expert who does not have any or little *methodological understanding* but the output of the system matches the expectations. While the user might have trust in this system, it gives the modeling expert great powers – and responsibilities. In a pessimistic

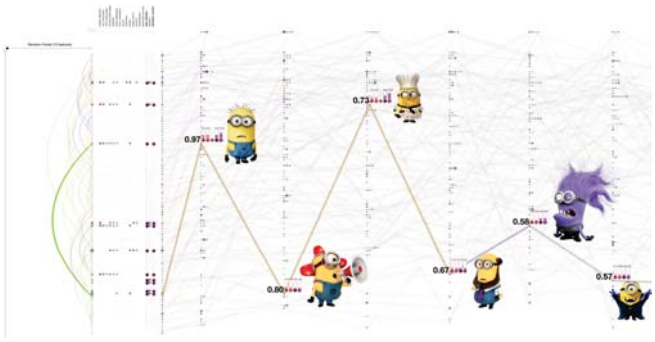

Fig. 5. *ThreadReconstructor* [5] as imagined by the study participants.

perspective also the great ability to manipulate the user. This situation is, however, not uncommon as we can experience this in many commercial products of our everyday lives, for example, in recommender systems of online-shops, search engines, and social networks. Such systems try to continuously adapt their output towards the user's expectations which imposes a high risk of including the user's biases and not producing objectively correct results. The consequence of this phenomena is also called "filter bubble."

Quadrant 2 is the desired state as only here the user can effectively use the system as the underlying methods are understood, and the output of the system is valid from the user's point of view. We consider this as the quadrant with the highest trust in the system and where it is likely that here the best conditions exist to generate more knowledge and validate existing knowledge. This is possible by using different data where the expected output is little or unknown and by varying parameter settings. Ultimately, the user should have understood the limitations of the system and the underlying methods. Lee et al. name this state a calibrated trust [19].

We consider quadrant 1 as an intermediate state where the user has a high *methodological understanding* but the output of the system does not match the user's expectations. However, the trust in the system is likely to be high. The user is therefore possibly motivated to validate the used models and processes or even check the implementation for errors. We refer to this process as *debugging the system*. In the case of finding an error on the concept-, implementation-, or even data-level, the user is adapting the model ($M \rightarrow D$) and thus progressing towards quadrant 2. If no errors can be discovered the user might be even willing to adapt her expectations towards the output ($D \rightarrow M$). This is mainly due to the higher trust in the system as compared to the bottom of the chart.

An advancement from quadrant 4 to quadrant 3 is possible but probably not as efficient. In this case, the model might randomly change the output due to the random parameter settings set by the user or the model applies an active learning methodology which typically only gradually changes the output. As the trust is missing the user will not be as persistent in using the system. In general, we consider this transition to be slower than from quadrants 1 to 2.

We propose metaphorical narratives as a method to elevate the domain expert in her *methodological understanding*. In Figure 4 this would result in transitions from quadrant 4 to 1 or 3 to 2, respectively. We further argue that a movement as depicted by the red arrow (Figure 4) is ideal for two reasons. First, the domain expert can validate the methods and may discover that some applied AI models are not suitable for the given task. This is especially important in the earlier stages of the design study and helps to prevent the time-consuming development of systems that turn out to be ineffective in supporting the domain expert in her tasks. Second, the user might be willing to adapt her expectations ($D \rightarrow M$). We consider the second effect as an essential part of the knowledge generation process.

While a state depicted by quadrant 3 is not desirable for the analysis of data in a scientific manner, the metaphorical narratives can be used to transition to quadrant 2 (black arrow).

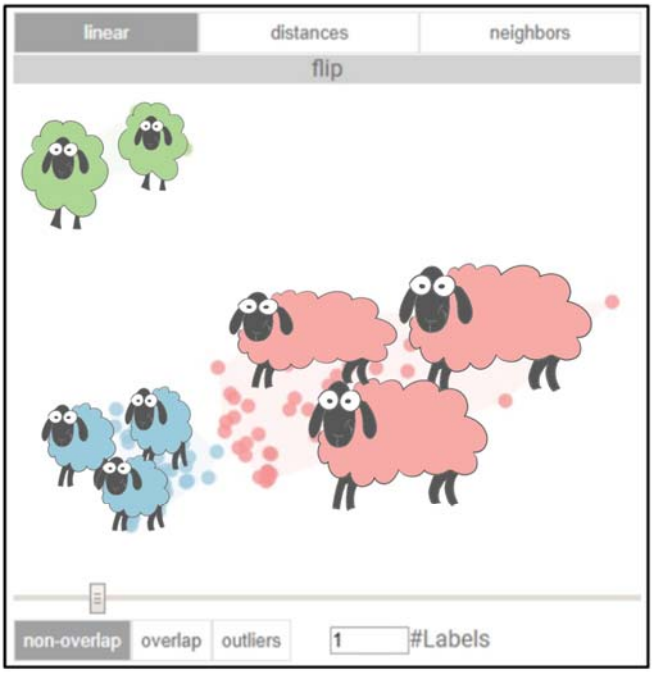

Fig. 6. The concept of a projection plot in combination with a visual clustering is explained by the metaphorical narrative of a sheep run.

3.3 Exemplary Metaphorical Narratives

In the following, we describe successfully used metaphorical narratives. In each case many alternatives are imaginable. However, a comparison of the effectiveness concerning the alternatives is not conducted.

Minions The *ThreadReconstructor* [5] was developed as a tool for the modeling of reply-chains to untangle conversations, e.g., as occurring in online forums and discussion sections. One essential property of this technique is the effective combination of pre-trained classifiers and user-defined queries based on some tailored features. To achieve an effective design for the targeted forum-analysts (with no prior knowledge of machine learning), this approach had to undergo a balancing act between simplicity and expressivity. Hence, a mapping of the decision space of all reconstruction models was developed. Here, the analysts could visually compare the performance of different models, as shown in Figure 5. During a pair-analytics session, however, the modeling, Visual Analytics Experts (VAE) noticed that using technical concepts, such as "classifier accuracy," were intimidating to the Subject Matter Experts (SME), as the latter feared to mess up with the configurations of the machine learning models. This prompted the VAE to spontaneously opt for a simplified explanation using the metaphorical narrative of *minions* [1] (think: features) operating in a magic-box (think: classifiers). According to this narrative, the sole aim of the minions (very specialized workers) was to optimize all circumstances to achieve the best quality for the one task they are good at (think: feature-optimization). Hence, within each magic-box, a different set of minions are operating to convert a tangled conversation into different threads. This metaphor was readily picked up by the SME and lead them to achieve comparable results in the study with a machine-learning expert. Converting the black box machine learning model into a magic-box with competing minions overcame the intimidating communication barrier and led to a successful deployment of *ThreadReconstructor*.

Sheep The *Concept Explorer* combines multiple complex AI models to support the criminal investigator in its Comparative Case Analysis task [14]. Two central AI models are dimensionality reduction techniques with weighted feature vectors and visual clustering techniques

[1] http://www.minionsmovie.com/, accessed July 2018

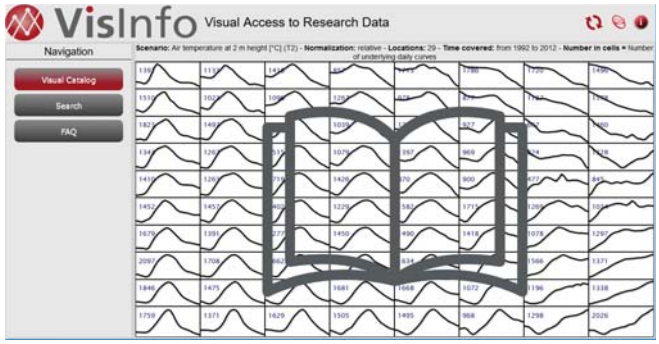

Fig. 7. The Visual Catalog of the VisInfo digital library system. The catalog narrative paraphrases the interactive browsing functionality, based on the complex SOM (Self-Organizing Maps) neural network algorithm.

that operate on the low-dimensional output of the dimensionality reduction model. To explain the difference between both methods and the general concepts behind them, we chose the metaphorical narrative of a flock of sheep (Figure 6). Sheep have different attributes such as size, length, and height. We explained that the domain expert can tell the shepherd what attributes she considers more or less important. The shepherd tries to place the sheep onto the sheep run based on how similar the sheep are according to their user-weighted attributes. Afterward, the user provides the shepherd with a set of colors. The shepherd tries to find groups of sheep on the sheep run without looking at their attributes and assigns each group one of the colors. The user can investigate and explore the groups, look at the distinctive attributes or find attributes that are shared among different groups. After teaching the basic concepts of dimensionality reduction and clustering techniques, the domain experts started to use the tool with much more confidence. The evaluations after establishing the metaphorical narrative showed that the users ceased their wishes for more guidance by the tool and observations confirmed the now more exploratory data analysis with the support of the system.

Fruits In a different project, working on digitizing the workflow of humanities scholars [26], we encountered the challenge of different scientific cultures having too diverse approaches to target a problem. While our colleagues from linguistics were concerned with the exact description of the different data points of a table, we as computer scientists wanted to identify commonalities in the data to define properties, categories, and most importantly, the dimensions of the data. After multiple unsuccessful attempts to instigate the linguists to think of their work in the abstraction of a data scientist, we opted for a simpler explanation. The narrative used to achieve consensus was that of a *fruit basket*. We challenged our colleagues from linguistics to describe the geometric attributes that make-up fruits, arriving at a listing of colors, shapes, etc. We then arranged the different fruits as rows of a table and these common attributes as column-headers. This simple toy-example was the missing bridge required to stimulate a common understanding between the parties involved in this project.

The Visual Catalog *VisInfo* is an exploratory search system for non-textual digital data content, created in a design study setting together with experts in the digital library domain [3]. In a very early stage of the domain and problem characterization phase, one collaborator came up with the metaphorical narrative of a *Visual Catalog*, showing important data content in a structured and intuitive way [2]. Users will be able to "browse" through the catalog, seeking interesting content for detailed analysis and downstream digital library support. The metaphorical narrative helped the user in understanding what we refer to as a data exploration task. Our implementation of the Visual Catalog was based on a SOM [17] (Self-organizing Map) neural network algorithm that combines vector quantization, dimensionality reduction, and (visual) clustering. Cluster visualization and interaction designs enable the exploration of the data content as well as query-by-example functionality. Given the visual representation of the model

output reflecting the notion of the Visual Catalog, we were able to hide details about the complex algorithm and algorithm parameters, as well as domain-specific terminologies such as clustering or dimensionality reduction. With the Visual Catalog, the principal behavior of our model was pre-defined, implementation details remained in the responsibility of our group. According to the Visual Catalog narrative, the digital librarians also had the means to understand both the principle concept of the SOM output (overview of the data in a structured/ordered way) as well as the interaction functionality (browsing, detail-on-demand). Just as well, the Visual Catalog was part of the validation strategy of the design study, allowing the digital librarians to assess whether we built the right product [23]. This example differs from the others as a metaphorical narrative is not used to describe the inner workings of the main idea of the system but helps the domain expert in understanding a crucial task that our tool supports. Additionally, the early on established metaphorical narrative impacted us in the design of the graphical user interface resulting that the user better understands *how* the *VisInfo* tool supports her in accomplishing this task. We choose this example to underline the wide applicabilities of metaphorical narratives.

4 Discussion

Metaphorical narratives have proven to be effective in our experience by explaining concepts of complex AI models to the domain experts in an intuitive way. This is also true vice-versa when domain experts explain concepts about their data, tasks, and domain problems.

Simplification We argue that the use of a metaphorical narrative encourages both sides to use an abstract, limited, and simplified vocabulary. This allows an easier transfer of the concepts while forcing the domain expert as well as the modeling expert to reduce details. Within their own field of expertise (and vocabulary), people focus much more on details and correctness of the concepts without asking themselves whether this level of detail is necessary for the other party to perform their respective tasks. The "curse of knowledge" may even lead the expert to the false conclusion that some fundamental concept is known to the other party. These two effects often result in the resignation of the other party as too much content is provided and many parts of the vocabulary are incomprehensible. A well-chosen metaphorical narrative leads to a focus on the main concepts of the used methods.

Visualization The metaphorical narratives have shown to be more effective when they include the visualizations that are available in the system. We described the example of a "sheep run as seen from the top with colored sheep on it" that fits well to the metaphor of a scatterplot (Figure 6). Views are meant to provide access for the domain expert to the underlying AI models and their output. Using and describing this access within the metaphorical narrative lowers the inhibition threshold for the user to use the system and the visual interface may become more usable and useful. This is especially important for the effective usage in data exploration.

Roles Furthermore, metaphorical narratives reduce the distance of the roles, i.e., the modeling expert role as "master of the data analysis methods" vs. the domain expert role as "master of the data and tasks". This difference in the roles lets the respective party hesitating to ask questions because of admitting to not understand a specific word or concept which is naturally clear to the other party. Using a unified and simplified vocabulary brings both parties closer together, allows to discuss at eye level, and encourages to ask questions.

Social Implications Metaphorical narratives may also have social implications. Funny or absurd metaphorical narratives can increase the motivation to bridge the gap from the own mental model to the metaphorical narrative and thus encourages an active discussion. This is, however, also risky as too much absurdness may result that the other party does not feel taken seriously. Other pitfalls include cultural, personal, or religious aspects that may offend the other party.

Abstraction Level The targeted level of detail impacts the choice of the metaphorical narrative greatly. The abstraction level has to be adjusted to the domain expert and the respective tasks. The metaphorical narrative should be expandable to add more details. It is possible to replace certain words with their respective domain-specific pendants as the user is memorizing the mapping.

Transitive Mapping A correct transitive mapping of the concepts of the respective domains through the metaphorical narrative (Figure 3) directly impacts the success or failure of the metaphorical narrative. As each party mentally performs this mapping towards their respective domain itself, it can be difficult to validate the other party's mapping. The vocabulary of the metaphorical narrative must, therefore, be chosen in a way such that each concept in the metaphorical narrative has no ambiguities. In our experience, this is mostly true for very simplistic metaphorical narratives where all used words of the vocabulary can be expected to be common knowledge. With the so established common ground, it can be better evaluated and communicated whether the mapping of the other party is correct or not than with a direct mapping of the domain-specific concepts. If one of the parties has a basic understanding of the concepts of the other party, it can further support and validate the correct mapping and may ease the overall process.

Success Indicators A good indicator for a successfully applied metaphorical narrative is when domain experts stick to the vocabulary throughout later phases of the collaboration. Especially, when discussing results or findings among each other but also when asking questions to the modeling experts. In our experience this can last for months if not longer. Additionally, domain experts exchange domain-specific words or terms from the methods with the vocabulary from the metaphorical narrative. Often times, this happened accidentally or without realizing it. A similar behavior can be observed with bilingual people. It shows that the mapping from the domain-specific terms to the rather technical terms of either domain via the transition of the metaphorical narrative was memorized very well. It is noteworthy that the process of exchanging domain-specific terminology with the vocabulary of the metaphorical narrative should not be steered or enforced by the creator of the metaphorical narrative. In a case where more details are needed, it is more effective to extend the metaphorical narrative itself. In our experience, it is inevitable that domain and modeling experts accidentally fall back to their own terminology. However, this might also have the positive side effect that the other respective party gradually understands the mapping of the metaphorical narrative to the domain of the other party.

Datafication The use of toy example datasets where the properties and effects in combination with AI models are known is a well-established practice, especially in the machine learning domain. We argue that a datafication of a metaphorical narrative can positively impact the effect of the metaphorical narrative as the output and the related visuals can be directly shown and explained in the system. It is therefore important to keep the dataset as simple as the metaphorical narrative itself such that the dataset can be easily inspected manually. The primary use of the dataset is not to show all possible effects that may occur in the system but to underline the inner workings of the system in combination with the metaphorical narrative.

5 Conclusions

In this work, we contribute the methodology of metaphorical narratives to achieve a mutual understanding of a common subject using abstract, simplistic, and visual vocabulary targeted at machine learning/AI. We argue that this methodology is complementary and well suited in combination with other established methodologies such as simplified model building. We ground our work on several disciplines, i.e., translation theory, didactics, software development, and state of the art design study methodologies. Additionally, we introduce a trust-building model that illustrates the benefits of our methodology. The field of visual analytics serves as a practical tool for corresponding implementations. In that context, we present four different metaphorical narratives that have been successfully implemented in interdisciplinary research projects to explain AI. Finally, we discuss different lessons learned as well as future work, related to important aspects of the proposed methodology.

REFERENCES

[1] S. Amershi, M. Cakmak, W. B. Knox, and T. Kulesza. Power to the people: The role of humans in interactive machine learning. *AI Magazine*, 35(4):105–120, 2014.

[2] J. Bernard, J. Brase, D. W. Fellner, O. Koepler, J. Kohlhammer, T. Ruppert, T. Schreck, and I. Sens. A visual digital library approach for time-oriented scientific primary data. *Int. J. on Digital Libraries*, 11(2):111–123, 2010. doi: 10.1007/s00799-011-0072-x

[3] J. Bernard, D. Daberkow, D. W. Fellner, K. Fischer, O. Koepler, J. Kohlhammer, M. Runnwerth, T. Ruppert, T. Schreck, and I. Sens. Visinfo: a digital library system for time series research data based on exploratory search - a user-centered design approach. *Int. J. on Digital Libraries*, 16(1):37–59, 2015.

[4] A. Chesterman. *Memes of translation: The spread of ideas in translation theory*, vol. 123. John Benjamins Publishing Company, 2016.

[5] M. El-Assady, R. Sevastjanova, D. A. Keim, and C. Collins. ThreadReconstructor: Modeling Reply-Chains to Untangle Conversational Text through Visual Analytics. *Comput. Graph. Forum*, 2018.

[6] G. Ellis. Cognitive biases in visualizations. 2018.

[7] G. Futschek. Extreme didactic reduction in computational thinking education. In *X World Conf. on Computers in Education*, pp. 1–6, 2013.

[8] M. Gleicher. Explainers: Expert explorations with crafted projections. *IEEE Trans. Vis. Comput. Graph.*, 19(12):2042–2051, 2013.

[9] G. Grüner. Die didaktische reduktion als kernstück der didaktik. *Die Deutsche Schule*, 59(7/8):414–430, 1967.

[10] C. Heath and D. Heath. The curse of knowledge. *Harvard Business Review*, 84(12):20–23, 2006.

[11] U. Hinrichs, M. El-Assady, A. Bradley, S. Forlini, and C. Collins. Risk the drift! stretching disciplinary boundaries through critical collaborations between the humanities and visualization. In *2nd IEEE VIS Workshop on Visualization for the Digital Humanities as part of the IEEE VIS 2017*, 2017.

[12] D. Hull, S. Pettifer, and D. B. Kell. Defrosting the digital library: Bibliographic tools for the next generation web. *PLoS Computational Biology*, 4(10), 2008.

[13] A. K. Jain. Data clustering: 50 years beyond k-means. *Pattern Recognition Letters*, 31(8):651–666, 2010.

[14] W. Jentner, D. Sacha, F. Stoffel, G. Ellis, L. Zhang, and D. A. Keim. Making Machine Intelligence Less Scary for Criminal Analysts: Reflections on Designing a Visual Comparative Case Analysis Tool. *The Visual Computer Journal*, 2018.

[15] D. A. Keim, J. Kohlhammer, G. P. Ellis, and F. Mansmann. *Mastering the Information Age - Solving Problems with Visual Analytics*. Eurographics Association, 2010.

[16] R. M. Kirby and M. Meyer. Visualization collaborations: What works and why. *IEEE Comput. Graph. and Applications*, 33(6):82–88, Nov 2013.

[17] T. Kohonen. The self-organizing map. *Neurocomputing*, 21(1-3):1–6, 1998.

[18] Y. LeCun, Y. Bengio, and G. Hinton. Deep learning. *Nature*, 521(7553):436–444, 2015.

[19] J. D. Lee and K. A. See. Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1):50–80, 2004.

[20] B. Meyer. *Object-Oriented Software Construction, 2nd Edition*. Prentice-Hall, 1997.

[21] R. S. Michalski, J. G. Carbonell, and T. M. Mitchell. *Machine learning: An artificial intelligence approach*. Springer Science & Business Media, 2013.

[22] T. Mühlbacher and H. Piringer. A partition-based framework for building and validating regression models. 19(12):1962–1971, 2013.

[23] T. Munzner. A nested model for visualization design and validation. *IEEE Trans. Vis. and Comput. Graph.*, 15(6):921–928, Nov. 2009.

[24] D. Sacha, A. Stoffel, F. Stoffel, B. C. Kwon, G. P. Ellis, and D. A. Keim. Knowledge generation model for visual analytics. *IEEE Trans. Vis. Comput. Graph.*, 20(12):1604–1613, 2014.

[25] D. Sacha, L. Zhang, M. Sedlmair, J. A. Lee, J. Peltonen, D. Weiskopf, S. C. North, and D. A. Keim. Visual Interaction with Dimensionality Reduction: A Structured Literature Analysis. 23(01):241–250, 2016.

[26] C. Schätzle, M. Hund, F. Dennig, M. Butt, and D. Keim. Histobankvis: Detecting language change via data visualization. In *Proceedings of the NoDaLiDa 2017 Workshop on Processing Historical Language*, pp. 32–39, 2017.

[27] M. Sedlmair, M. Meyer, and T. Munzner. Design study methodology: Reflections from the trenches and the stacks. *IEEE Trans. Vis. Comput. Graph.*, 18(12):2431–2440, Dec 2012.

[28] S. Simon, S. Mittelstdt, D. A. Keim, and M. Sedlmair. Bridging the Gap of Domain and Visualization Experts with a Liaison. In E. Bertini, J. Kennedy, and E. Puppo, eds., *Eurographics Conf. on Visualization (EuroVis) - Short Papers*. The Eurographics Association, 2015. doi: 10.2312/eurovisshort.20151137

[29] C. Tominski, J. F. Donges, and T. Nocke. Information visualization in climate research. In *15th Int. Conf. on Information Visualisation, IV 2011, London, United Kingdom, July 13-15, 2011*, pp. 298–305, 2011.

[30] J. Tulach. *Practical API Design: Confessions of a Java Framework Architect*. Apress, Berkely, CA, USA, 1 ed., 2008.

[31] J. J. V. Wijk. Bridging the gaps. *IEEE Comput. Graph. and Applications*, 26(6):6–9, Nov.-Dec. 2009.