SECTION 2

Decomposing the explanation process

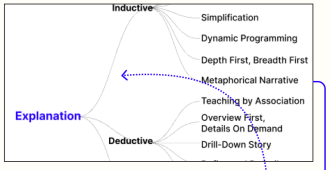

Fig. 2: Hierarchical arrangement for components

ID	Medium	Explanation
015	Visualisation, Images	Comparisc
016	Visualisation	Depth Firs
017	Visualisation, Images	Drill-Down
018	Visualisation, Images	Drill-Down
019	Visualisation	Drill-Down
020	Images, Game	Drill-Down
021	Visualisation, Game, Image	Overview F

Fig. 3: Coding of case studies according to components

SECTION 3

The XAI Primer

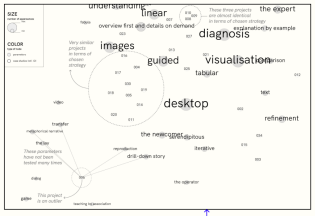

Fig. 7: Visualisation study

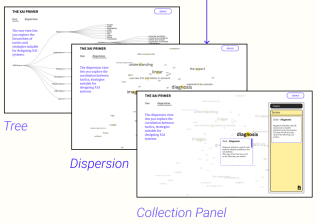

Fig. 4-6: Main views

Figure 1: Overview of this paper, including the main figures.

The XAI Primer: A Digital Ideation Space for Explainable Artificial Intelligence Strategies

Beatrice Gobbo
Politecnico di Milano
Milan, Italy
beatrice.gobbo@polimi.it

Mennatallah El-Assady
Konstanz University
Germany
menna.el-assady@uni.kn

Components of Explainable Artificial Intelligence Processes.

Building Blocks

- **Explanation**: the type of reasoning strategy;
- **Verification**: the strategy for ensuring users understood the subject;

Building Block Attributes

- **Task**: the action to be carried out through the explanation;
- **Data**: the type of data employed in the system;
- **Medium**: the combination of media and language adopted;
- **Path**: the way in which blocks are connected;
- **Exploration**: the type of navigation and exploration allowed;
- **User**: the target user;
- **Scenario**: the usage and fruition scenarios.

Abstract

Explainable Artificial Intelligence (XAI) processes typically combine various explanation and verification strategies to support the analysis in different domains. Due to the increasing number of techniques and the variety of XAI methods deployed, deriving a comprehensive overview framework of different strategy combinations remains challenging. The paper presents a proposal for a digital ideation space in which designers of XAI processes can derive inspiration and investigate existing works. We propose an exploratory interface depicting both XAI strategies and applications to support designers conceptualizing and developing new projects. The *XAI Primer* is designed based on the metaphor of a museum. Users can explore the presented ideation space as if they were artists visiting an art gallery. We enable serendipitous and guided explorations, allowing them to investigate and probe the state-of-the-art as a source of inspiration.

Author Keywords

Explainable Artificial Intelligence; Explainable Artificial Intelligence Guidance; Data Visualisation; Serendipitous Exploration

CCS Concepts

•Human-centered computing → visualisation theory, concepts and paradigms;

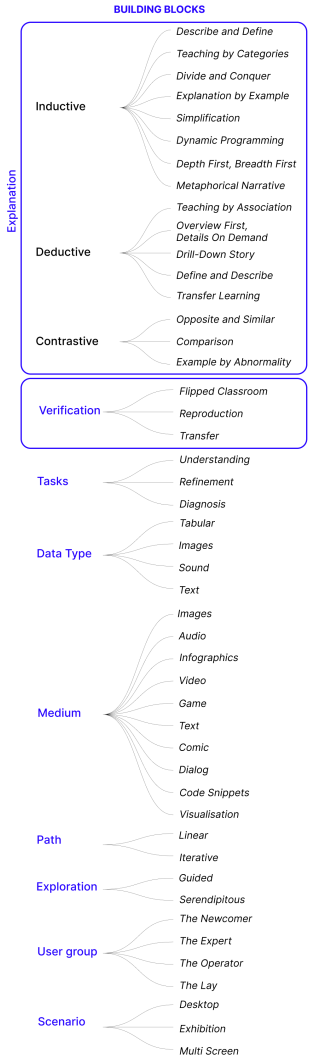

Figure 2: A hierarchical synopsis of **XAI strategies** [4].

Introduction

Explainable Artificial Intelligence relates to the research and application of methods for facilitating the understanding of fully automated decision-making processes. The General Protection Data Regulation [1] highlighted the need to provide stakeholders affected by opaque AI-driven systems with the tools to interpret automated decision-making processes. Thus, to fulfill this need, scholars from diverse backgrounds have designed techniques and interfaces to enable the interpretation of AI-based decisions. These approaches tackle the explainability challenge at different degrees of sophistication. Recent surveys have grouped the proposed works into several categories and classes [5, 8, 10]. Additionally, efforts have been made to structure the explanation process into conceptual frameworks., for instance, based on the study of strategies and best practices coming from different research domains (including *pedagogy*, *gamification* and *storytelling*) [4]. Furthermore, a recent study focused on building a prototyping tool to design Explainable Artificial Intelligence for non-technical end-users, combining techniques specific to the field of data visualisation [9].

Design Motivation

Given the still young nature of the research field and its rapid and constant growth, it's necessary to provide XAI designers (both designers of the system and designers of the interface) with a comprehensive and collaborative *space* for exploring the state-of-the-art of the field and for boosting experimentation with approaches not yet explored. For that purpose, in this paper (Figure 1), we introduce the design considerations of our proposed ideation space. The XAI Primer will be developed to enable users to preview existing cases, as well as to contemplate new strategic combinations [3]. Furthermore, to strengthen our approach's interactive and exploratory nature, we propose a space where the visual and metaphorical aspects are primary concerns of the design.

Decomposing Explanation Processes

We introduce the XAI Primer as a visual space to explore works based on the dimensions proposed as the building blocks of explanation processes [4]. The essential conceptual elements of an explanation process are listed on the first page of this paper and schematically depicted in Figure 2. Three **tasks** have been identified as prominent for XAI: *understanding*, *diagnosis* and *refinement*. These define the main **phases** of an XAI process, which are further subdivided into **building blocks** consisting of explanation and verification strategies. **Explanations** are classified according to their strategy as *inductive, deductive* or *contrastive*; while **verifications** are based on strategies for *reproducing* or *transferring* knowledge, or on *flipped classrooms*. Building blocks could be hinged and connected *linearly* or in an *iterative* way, as well as through *guided* or *serendipitous* exploration [4].

Target **user groups** can have different levels of expertise, which implies tackling various degrees of complexity when providing an explanation [7]. Additionally, **data** types analyzed are usually variable. Hence, to tailor the XAI process to the data and users, various kinds of **mediums** can be used during explanation and verification [4]. Finally, the **usage scenario** must be considered (for instance, whether it is a desktop application or an interactive museum installation). We rely on these dimensions to encode XAI approaches (see Figure 3), and to span the design space of the XAI Primer.

Case Study ID | Parameters

ID	Medium	Explanation	Pathaway	Joint	Verification	Tasks	Data Type	User Group	Scenario
015	Visualisation, Images	Comparisc	Guided	Linear		Understa	Images	The Oper	Desktop
016	Visualisation	Depth Firs	Serendipi	Iterati	Reproduc	Understa	Images	The Oper	Desktop
017	Visualisation, Images	Drill-Down	Guided	Linear	Transfer	Understa	Tabular	The Newc	Desktop
018	Visualisation, Images	Drill-Down	Guided	Linear	Transfer	Understa	Images	The Oper	Desktop
019	Visualisation	Drill-Down	Guided	Linear	Reproduc	Understa	Images	The Newc	Desktop
020	Images, Game	Drill-Down	Guided	Linear	Transfer	Understa	Images	The Lay	Desktop
021	Visualisation, Game, Image	Overview I	Guided	Linear	Transfer	Understa	Images	The Newc	Desktop

Figure 3: Excerpt from matrix of case studies and XAI strategies. So far, we classified 35 approaches from the VISxAI Workshops.

Figure 4: The tree view offers a hierarchical synopsis of the components.

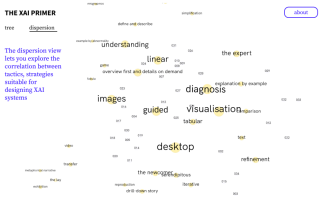

Figure 5: The projection perspective offers an overview of relations between strategies and their application in real projects. It uses MDS and a Force-Directed Spatialisation.

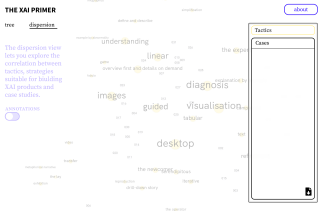

Figure 6: The collection panel is the space where users can gather suitable elements for the ideation stage in the design process.

The XAI Primer

The Explainable Artificial Intelligence Primer is an interactive hornbook (a book that serves as primer for study) to help designers conceive suitable XAI strategies. The exploratory interface is composed by the *tree view* (Figure 4) and a *projection view* (Figure 5). Moreover, the latter is equipped with a superimposed *collection panel* (Figure 6).

Data Coding

We designed the XAI Primer's exploratory space to consist of two types of objects; the previously mentioned strategies and artefacts encoding case studies for XAI. Thus, we first collected a preliminary sample of recent works from the *VISXAI Workshops* as case studies and then categorized them according to the previously mentioned dimensions. While some cases were easily attributable to categories, it was not entirely evident or simple to find a suitable location for others. For instance, in the *Théo Guesser* project, depicted in Figure 3, it is straightforward to identify the scheme (i.e., a clear combination of employed mediums and the inclination to present *overview first*, followed by *details on demand*).

Backbones and Visual Design Rationale

Since our goal is to provide designers with an exploratory space, the first step was to devise an interface and populate it. The sample of case studies and related strategies previously described have been encoded as visual objects. Using Multi-Dimensional Scaling (MDS), we project them on the canvas, thus depicting their similarity through their position's proximity. This basic layout aims to reveal the relationship between explanation strategies and their application on real projects. The drafted visualization is the digital space's backbone where both, case studies and strategies, can be positioned. Figure 7 shows a view of the XAI Primer.

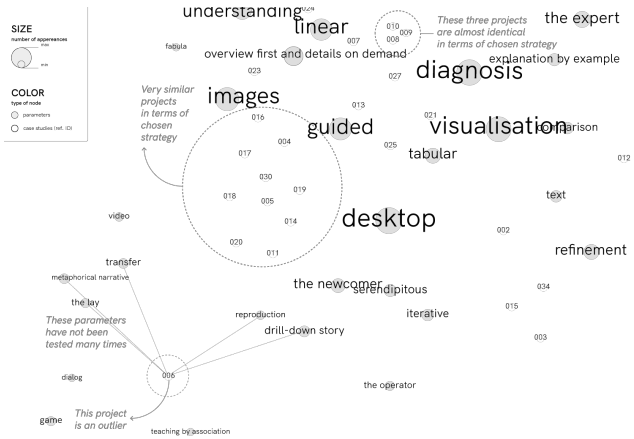

Figure 7: Study of the main visualisation, built upon a combination of MDS and a Force-Directed spatialisation. Annotations, provided by us, highlight clusters and void parts of the space.

An Early Glimpse of the Interface

The XAI Primer is an interactive space where designers find inspiration and guidance to compose their strategy and process. As mentioned before, the main view is the *projection view* (Figure 5) whose purpose is to display items clustered according to their similarity and to highlight portions of the space at various crowding stages. Users can also switch to the *tree view* (Figure 4) to access a structured and hierarchical visual taxonomy of the available strategies. Users can access detailed information about the items by hovering them, including parent relation indicators, descriptions for strategies, project summaries, and links to related publications (Figure 8). Besides, a superimposed *panel* is designed to give the users the chance to collect and combine items, divided into case studies and strategies and with guidance on their mutual relations. Finally, at the end of the exploration, users can download the content of the panel, thus, acquiring the blueprint of their collection.

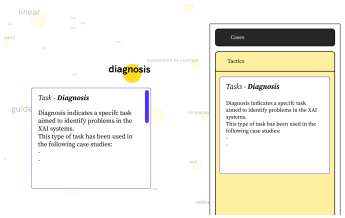

Figure 8: By hovering elements, users can access additional information.

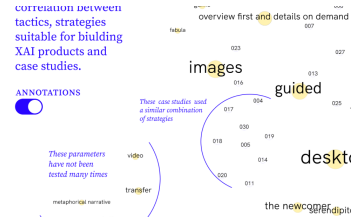

Figure 9: A zoom of the annotation layer providing additional information about arrangements.

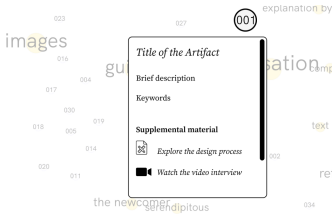

Figure 10: The future embedding of additional material to each artifact would ensure a deeper and more insightful reading.

Navigating the Space of the XAI Primer

The XAI Primer has no finite spatial boundaries; instead, we aim to make it a collaborative tool that users can expand by submitting additional case studies and any new strategy. This mixture of information from various disciplines and different methodological approaches can enrich the virtual space generated by a preliminary selection of items.

Serendipitous Exploration

Users could be imagined as serendipitous visitors of that digital space, moving through a growing interface populated by XAI case studies and strategies. The purpose is to provide a space that can be explored as if users were surrounded by *artefacts* [6] in a museum setting where both casual and guided exploration is envisioned.

Guided Tours

In order to leave room for serendipity and inspiration, the exploration is not intended to be predominantly guided: it will be at the user's will to turn on or off a level of annotations that highlights a critical reading of the space (Figure 8). Annotations, widely used in digital humanities, provide tutored and critical readings of the items' relationships and positioning. For instance, on the one hand, from the sample visualisation depicted in Figure 9, a group of floating strategies is evident and, looking in detail, it is possible to discern among them *lay user, metaphorical narratives* and *exhibition*: a combination of strategies (Figure 9a) that, within the sample of artefacts that we selected to develop the concept, was used only once (Figure 9b) [11]. On the other hand, in the central area, the space is crowded of artefacts that present *desktop* applications, with *images* and *visualisation*, mainly designed to complete a *diagnosis* task, following a *linear* path with a *guided* exploration. Retaining the museum metaphor, we envision crafting several tours based on the users' interests and expertise.

Research Opportunities

This paper proposes the operationalisation of a conceptual framework based on the combination of XAI strategies. However, to strengthen the exploratory space, it is necessary to increase our data collection, highlight the saturated areas, and open gaps in the space.

Enriching the Displayed Artefacts

We envision extending the displayed artefacts by detailed descriptions and interviews provided by authors of existing systems. For example, the ideation space can contain videos and supplementary material on the design process (Fig. 10).

Ideation Seekers

Our serendipitous navigation facilitates the creative exploration process without providing any sticky entry point. The serendipitous visitor retraces the behaviour of the information seeker portrayed by M. Dörk et al. [3]. To enhance the role of *information seekers*, we envision giving users the option of downloading a blueprint, summarizing their exploratory path, and providing them with a communication tool.

Conclusion

We presented the XAI Primer, a visual interface for ideation based on XAI strategies and artefacts from existing works. This paper presented a proposal for a digital ideation space where XAI processes designers can find inspiration and explore the state-of-the-art. Thus, the XAI Primer is a conceptual attempt to adopt design-thinking methods (including *mindmap* and *brain walk*) [2] in XAI, exploiting visual tactics from the domains of digital humanities and information visualisation. In our future work, we aim to extend the ideas presented in this paper to a full system implementation, incorporating a larger use case collection from existing work.

REFERENCES

[1] 2016. European Parliament and Council of European Union. Regulation (EU) 2016/679 (2016). https://eur-lex.europa.eu/legal-content/EN/TXT/HTML/?uri=CELEX:32016R0679&from=EN

[2] Rikke Friis Dam and Yu Siang Teo. 2020. Stage 3 in the Design Thinking Process: Ideate. (2020). https://www.interaction-design.org/literature/article/stage-3-in-the-design-thinking-process-ideate

[3] Marian Dörk, Sheelagh Carpendale, and Carey Williamson. 2011. The information flaneur: a fresh look at information seeking. In *Proceedings of the International Conference on Human Factors in Computing Systems, CHI 2011, Vancouver, BC, Canada, May 7-12, 2011*, Desney S. Tan, Saleema Amershi, Bo Begole, Wendy A. Kellogg, and Manas Tungare (Eds.). ACM, 1215–1224. DOI: http://dx.doi.org/10.1145/1978942.1979124

[4] Mennatallah El-Assady, Wolfgang Jentner, Rebecca Kehlbeck, Udo Schlegel, Rita Sevastjanova, Fabian Sperrle, Thilo Spinner, and Daniel Keim. 2019. Towards XAI : Structuring the Processes of Explanations. In *Proc. of ACM CHI Workshop on Human-Centered Machine Learning*.

[5] Rafael Garcia, Alexandru C. Telea, Bruno Castro da Silva, Jim Tørresen, and João Luiz Dihl Comba. 2018. A task-and-technique centered survey on visual analytics for deep learning model engineering. *Computers and Graphics (Pergamon)* (2018). DOI: http://dx.doi.org/10.1016/j.cag.2018.09.018

[6] Beatrice Gobbo. 2020. Explaining AI Through Critical Reflection Artifacts - On the Role of Communication Design Within XAI. In *Advanced Visual Interfaces. AVI 2020 Workshops, AVI-BDA and ITAVIS,Ischia, Italy*, Vol. 12585. Springer, 184–188. DOI: http://dx.doi.org/10.1007/978-3-030-68007-7_12

[7] Riccardo Guidotti. 2020. Explaining Explanation Methods. In *Proceedings of the First Workshop on Bridging the Gap between Information Science, Information Retrieval and Data Science (BIRDS 2020) co-located with 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2020), Xi'an, China (online only), July 30th, - 2020 (CEUR Workshop Proceedings)*, Ingo Frommholz, Haiming Liu, and Massimo Melucci (Eds.), Vol. 2741. CEUR-WS.org, 6–13. http://ceur-ws.org/Vol-2741/paper-04.pdf

[8] Fred Hohman, Minsuk Kahng, Robert Pienta, and Duen Horng Chau. 2019. Visual Analytics in Deep Learning: An Interrogative Survey for the Next Frontiers. *IEEE Transactions on Visualization and Computer Graphics* 25 (2019), 2674–2693. DOI: http://dx.doi.org/10.1109/TVCG.2018.2843369

[9] Weina Jin, Jianyu Fan, D. Gromala, P. Pasquier, and G. Hamarneh. 2021. EUCA: A Practical Prototyping Framework towards End-User-Centered Explainable Artificial Intelligence. *ArXiv* abs/2102.02437 (2021).

[10] Yafeng Lu, Rolando Garcia, Brett Hansen, Michael Gleicher, and Ross Maciejewski. 2017. The State-of-the-Art in Predictive Visual Analytics. *Computer Graphics Forum* 36, 3 (2017), 539–562. DOI:http://dx.doi.org/10.1111/cgf.13210

[11] Mauro Martino, Hendrik Strobelt, Owen Cornec, and Evan Phibbs. 2018. Forma Fluens. (2018). http://formafluens.io/index.html#first

FIGURE 2

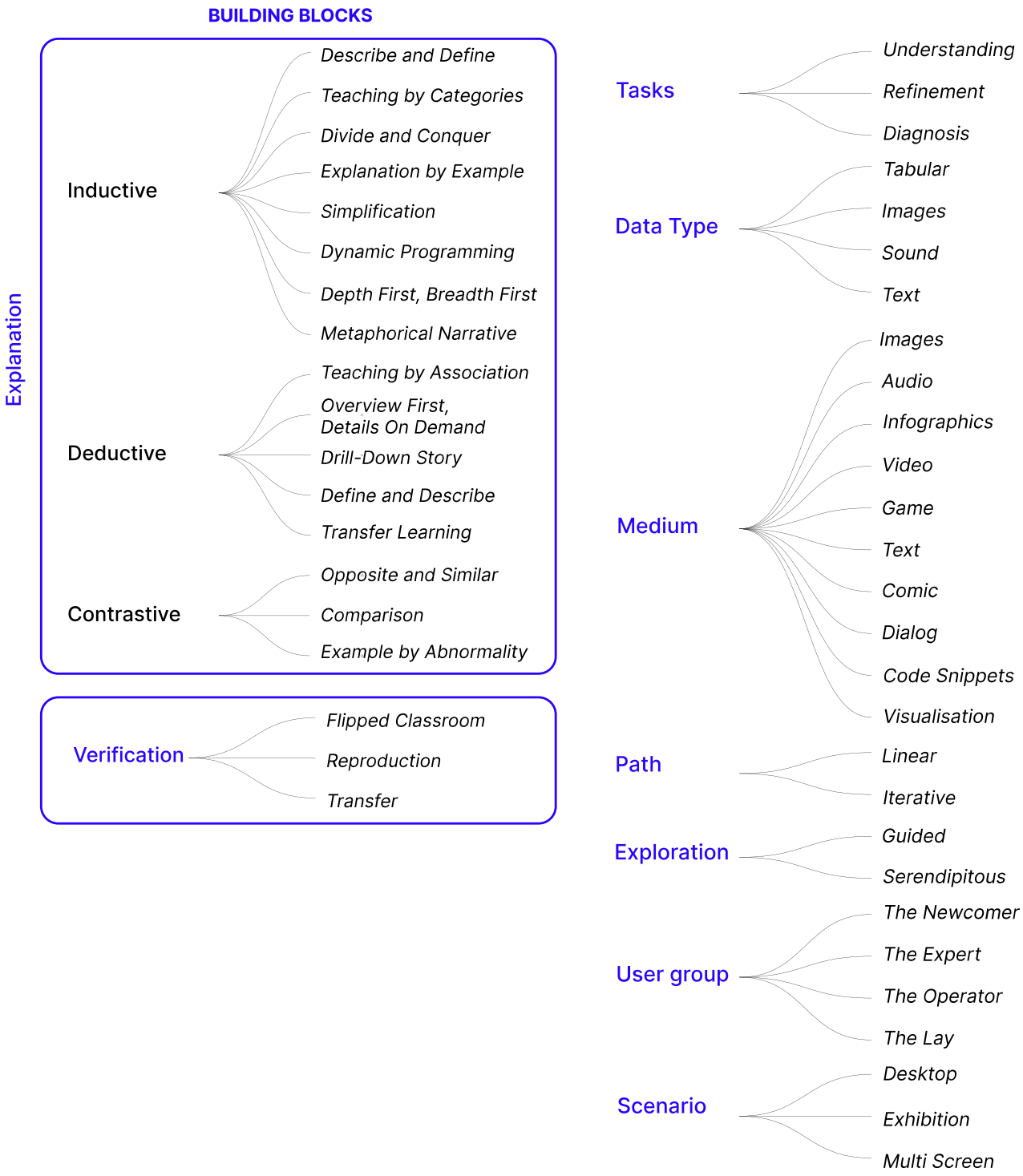
BUILDING BLOCKS
Explanation
Inductive
Describe and Define
Teaching by Categories
Divide and Conquer
Explanation by Example
Simplification
Dynamic Programming
Depth First, Breadth First
Metaphorical Narrative
Deductive
Teaching by Association
Overview First, Details On Demand
Drill-Down Story
Define and Describe
Transfer Learning
Contrastive
Opposite and Similar
Comparison
Example by Abnormality
Verification
Flipped Classroom
Reproduction
Transfer
Tasks
Understanding
Refinement
Diagnosis
Data Type
Tabular
Images
Sound
Text
Medium
Images
Audio
Infographics
Video
Game
Text
Comic
Dialog
Code Snippets
Visualisation
Path
Linear
Iterative
Exploration
Guided
Serendipitous
User group
The Newcomer
The Expert
The Operator
The Lay
Scenario
Desktop
Exhibition
Multi Screen

FIGURE 4

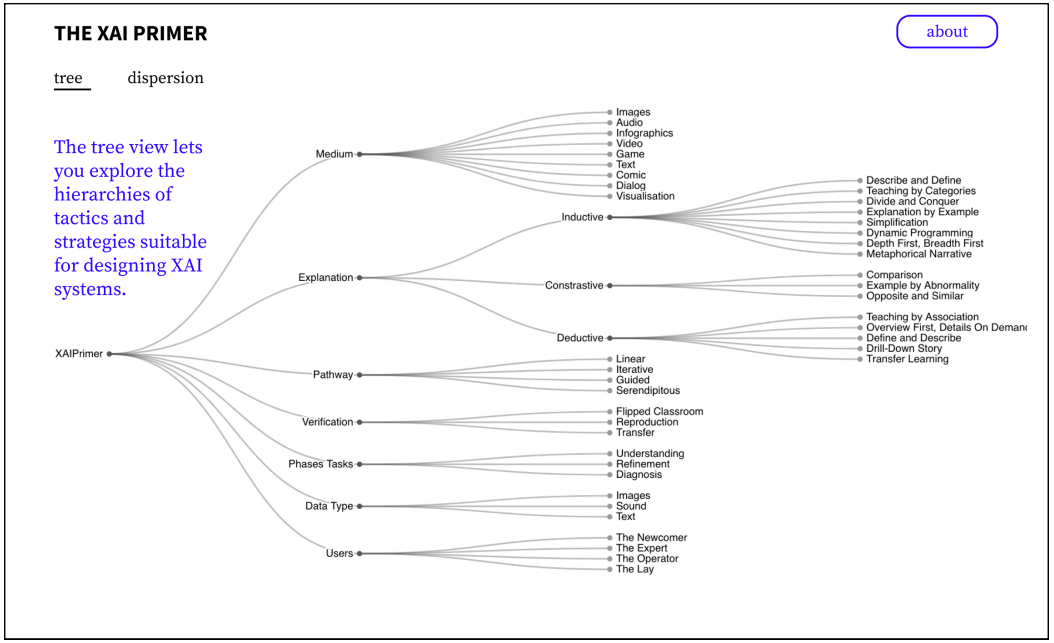

FIGURE 5

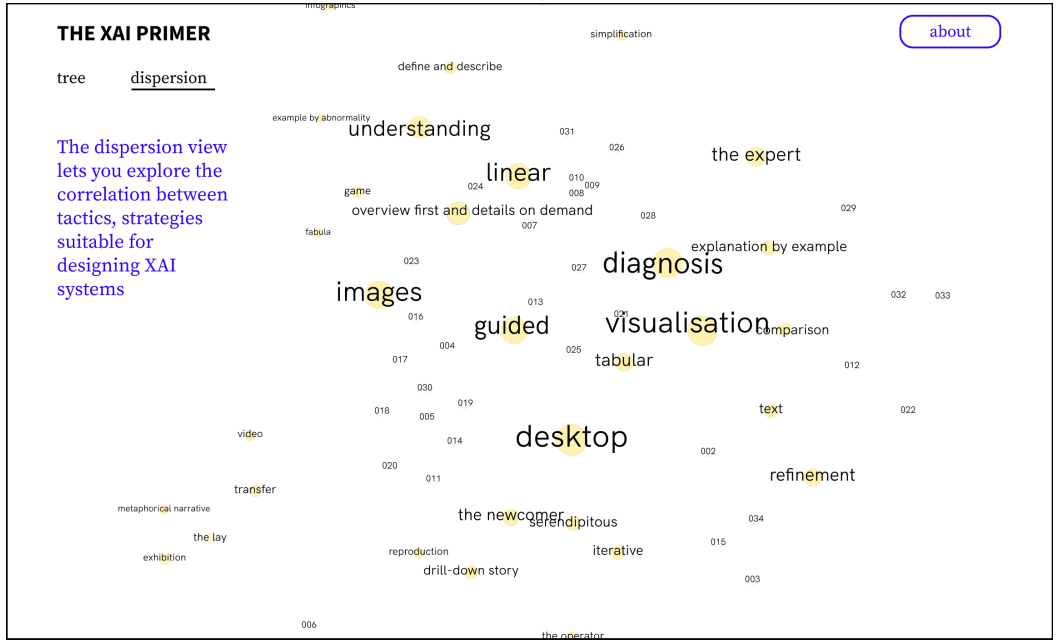

FIGURE 6

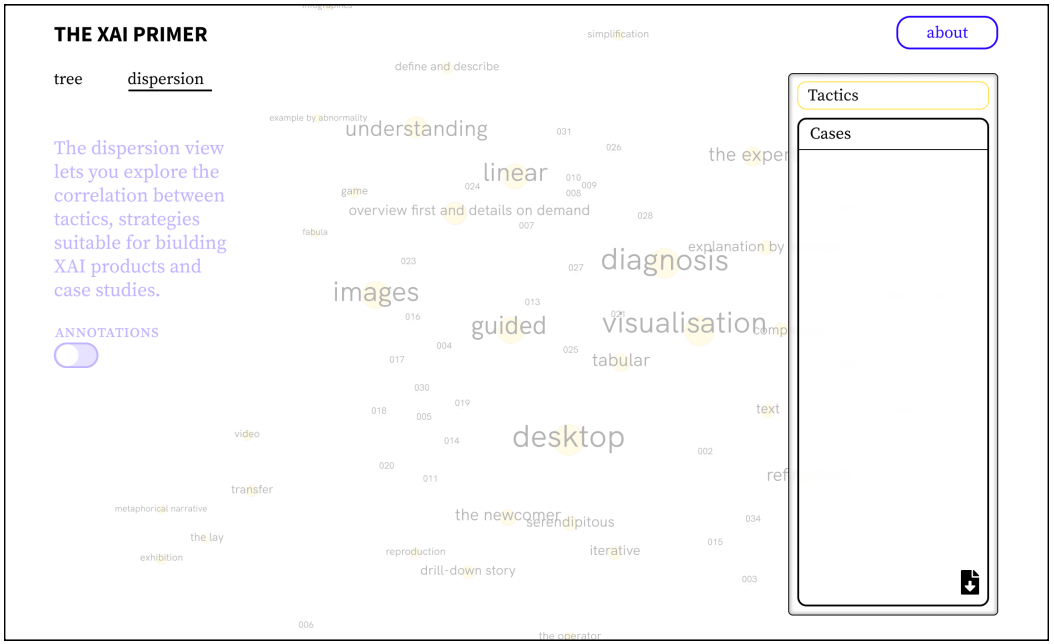

FIGURE 7

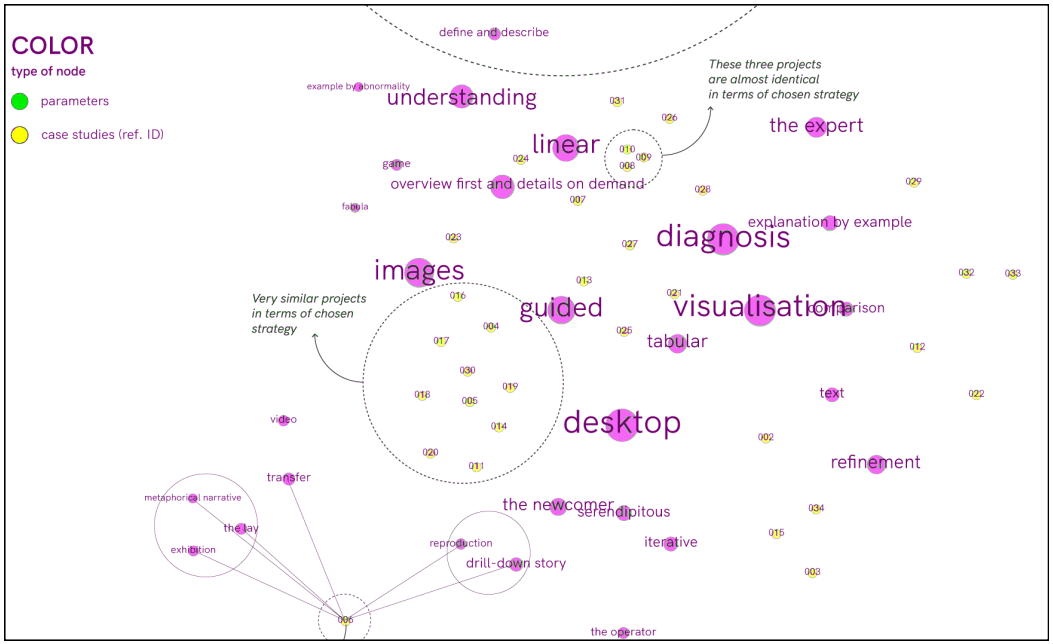

FIGURE 8

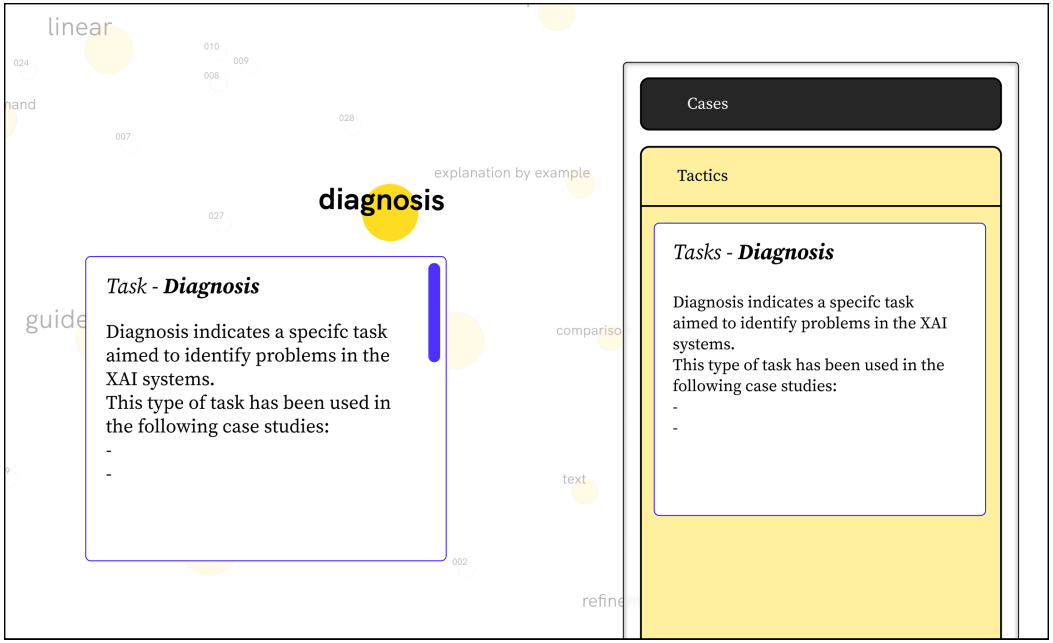

FIGURE 9

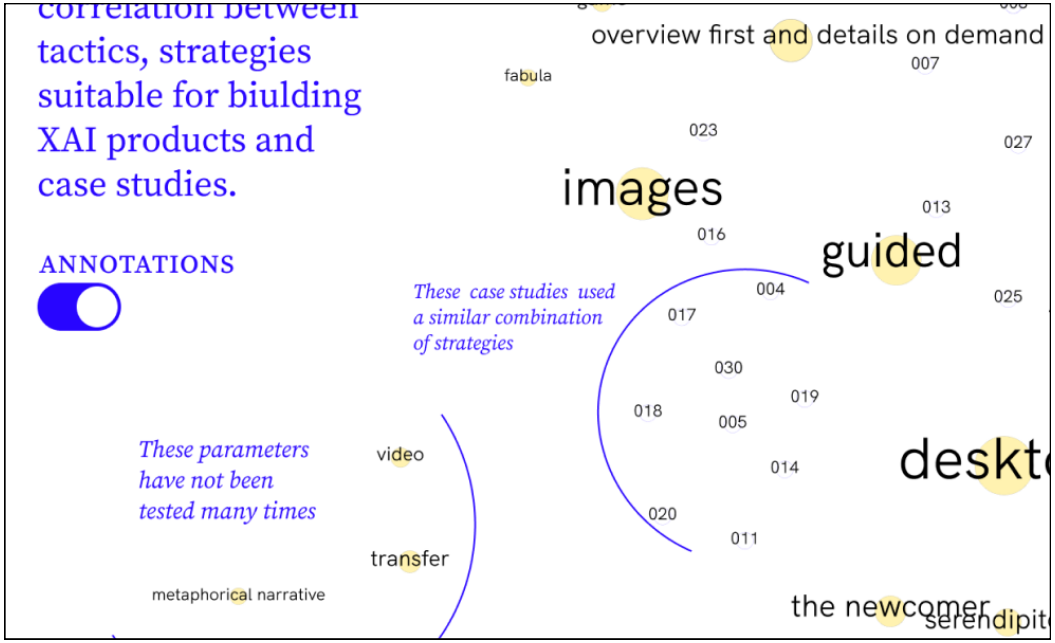

FIGURE 10