DATA SCIENCE HOLODECK

2023

Investigation and integration of Artificial Intelligence and Virtual Reality for exploration, comprehension, and transformation of complex data into usable knowledge

cphbusiness
COPENHAGEN BUSINESS ACADEMY

ABSTRACT

Nowadays, the innovative Information Technologies (IT) create numerous new opportunities for individuals and business users with a speed hardly reachable by humans implementing them for a benefit. The more complex the matter is, the steeper and longer its adaptation and accommodation process takes effect. One such complex family of technologies addresses the data-driven business and knowledge-based support of management decisions.

Human activities, such as learning, exploring, socializing, as well as business intention of knowing the customers and the market, product design and architecture, warehousing and logistic, optimization and value chain management, recognition of patterns and diagnostics of problems - all involve reference to data and require skills in abstraction, attention, identification, aggregation, and sometimes, deeper knowledge of mathematics, linguistics, and algorithmics*[*].

The project Data Science Holodeck aims at research and development of technologies that can support the cognitive activities mentioned above. In particular, we focus on democratization and implementation of Artificial Intelligence (AI) and Virtual Reality (VR) as powerful and insightful engines for revealing, visualization, and explanation of data-driven artefacts in higher education and smaller enterprises.

The project work and outcomes address the difficulties and the obstacles that keep numerous businesses and academic institutions away from the endless benefits unlocked by the emerging innovative technologies in support of knowledge-based human activities, which involve analytical reasoning and abstract thinking. The project provides proven in practice, workable, and affordable methodological instruments and software solutions accommodating Machine Learning (ML), Natural Language Processing (NLP), Knowledge Graphs (KG) and Mixed Reality (MR), supplemented by clear, readable, and understandable guides for their successful implementation in variety of data-driven business and learning practices.

Todorka Dimitrova, PhD

Project Manager

tdi@cphbusiness.dk

[*] Resistance Is futile**
[**] Star Track

INDEX

PROJECT FOUNDATION

Motivation

The Holodeck idea (originally from the above mentioned Star Track) came to us a couple of years ago with the first larger popularisation of big data, machine learning, and virtual reality games in education and business. Being qualified and experienced natural science and information technology practitioners, deeply involved in teaching and training ourselves, we were able to observe the difficulties in learning and implementation of the related abstract concepts, forming the foundation of algorithmics, data processing and analytics. Many of our students, colleagues and business partners were either seeing challenges in diving into the matter or lacking sufficient time and human resources to deal with it.

Data Science Holodeck offered innovative approach to support of humans and democratisation of technology.

It wasn't before Meta and the NLP AI (OpenAI's ChatGPT in particular) came into the scene, when the majority of individuals and SMEs (small-end-medium size enterprises) with non-technical background in Denmark started turning their attention to non-public data sources and planning to use the major ML and VR achievements with a purpose.

At present, the world witnesses a dramatic shift towards implementation of advanced IT. The tornado effect of the emerging generative AI-based technologies on the entire society, education and business communities confirms the contemporality and the relevance of our project ideas.

As the ideas are already implemented and outcomes are already available, the impact scale and benefits of Data Science Holodeck would be even larger than initially intended.

Objectives

The primary project objective has been creating a support for cognitive human activities and operations, related to working with data and complexity, such as driving attention, understanding, remembering, reasoning, and similar, by

- applying advanced AI methods and
- accommodating interactive VR technologies.

The project solution's building foundation lies on the assumption that involving all human senses in the process promotes human attraction to and understanding of complex problems and tasks with data, information extraction and knowledge acquisition.

The parallel secondary project objective has been the democratisation of the AI artefacts, disseminating them in an explainable and accessible way to individuals and small businesses, which are driven without direct support of data scientists and engineers.

The expectation is that the support for the preparation and taking informed decisions involving data and AI would contribute to the increase of the awareness, optimisation of operations' performance and improving the results in terms of shortening the data processing time and reducing the number of trials and failure.

Last, but not least, the project work has been directed to collecting, framing, and distributing facts and evidences about the innovative and still understudied concepts and pilot implementations of AI and VR into common business and education activities. Transferring the newly consolidated knowledge and expertise and making them widely available through Internet repositories, world-wide webs, and public events would help other research teams, educators, and data science practitioners to start higher and achieve results in broader range of aspects in the subject matter.

Target Groups

The objectives are designed for the benefit of the following target groups of users:

- SMEs developing IT and other businesses that operate with data
- higher education academics – students, educators, and study managers engaged in building the intelligent instruments and data-related power of modern IT implementations
- data scientists and knowledge managers, interested in creating of advanced AI and VR applications for common use in various professional and social implementation areas

Scope

Applying AI and VR in support of human decisions over data complexity is a large complex system of methods, tools and techniques itself. It is an interdisciplinary project involving integrated research and development in multiple directions: data science, computer technology and software development, human factors and psychology, business process management and decision-support systems, mathematics and statistics, machine learning and ontologies, graph theory, linguistic, and immersive visualization.

In the core focus of the research interest and the development work have been:

- using natural human language in data operations and automated document processing
- building and implementing knowledge graph as a domain context
- creating and experimenting with software instruments delivering AI and VR
- creating interactive visualization and visual explanation of data exploration and analysis
- prototyping and evaluating of 3D and VR interfaces and their impact on users' experience

Team

The project team consists of Todorka Dimitrova and Jon Bertelsen, lecturers at Copenhagen Business Academy.

Research and Development Methodology

The project work consists of selection, generation, and application of set of research and development methods, tools, and technologies. Representation and processing of complex data requires implementation of methods from mathematics, statistics, computer graphics and visualization, as well as ML and DL algorithmics.

Design Thinking

The project work has been organised around the Design Thinking methodology, gradually moving from the ambitious initial ideas toward advanced and complex developments and implementations.

Design thinking is an agile iterative process that allows the team to exchange ideas, regard assumptions, redefine the tasks and suggest alternative innovative solutions for prototyping and testing with the potential users.

We have staged the process in four phases and groups of activities (see https://innotechspace.dk/holodeck/methodology/

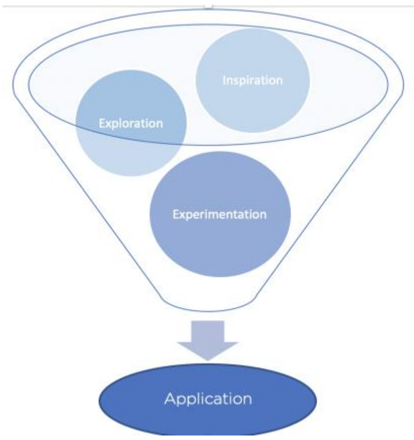

Figure 1 Design Thinking Phases Source: author

- Inspiration – research, analyse, empathize, identify, and define specifications
- Exploration – brainstorm, ideate, and prototype to validate and adjust the ideas
- Experimentation – carry experiments, test prototypes, and collect next level of feedback
- Application – create applications, deploy, and disseminate the results, prepare for Holodeck 2.0

Based on our previously accumulated knowledge, teaching and researching experience, in collaboration with students and partners from enterprises, we were able to identify some of the challenges the learners and business users experience, while accommodating and implementing complex data structures and processing algorithms.

Review of Resources

To discover more, as part of the pre-study, we undertook an extensive review of published resources and best practices in the interdisciplinary area, where big data, data science, and data-driven business meet computer technology, artificial intelligence, and immersive visualization (see https://innotechspace.dk/reference/).

Additionally, the streams of on-line bulletins `Medium` and `Medium Daily Digest` (https://medium.com), `KDnuggets` (https://kdnuggets.com), `The Variable - A newsletter by Towards Data Science`, `DigiTech` and `DataTech` from (https://ing.dk), as well as the AI section of `Arxiv` (https://arxiv.org) have been followed on a permanent basis during the recent years.

The review of publications has provided insights about the current state-of-the-art and the major trends in our field. It contributed to the definition and specifications of project goals, as well as to narrowing down the scope of the research and development initiative.

Some highlights of the review show:

- Quite a limited number of sources discuss the integration of AI and VR in applications. These two powerful technologies collaborate successfully in VR computer games, where AI is used for controlling characters' behaviour.

- During the recent couple of years VR technologies find broader implementation in business applications, as well as in teaching and training platforms. The VR/XR object modelling and event simulation ability matches the necessity of artificially created environments, where the trainees can undertake safe experiments and practice the development of new skills. Just a few commercial applications address the use of VR for data visualisation, and even less comment support of analytical tasks.

- There is a huge gap in the field of 3D and VR UI design. The problem has multiple aspects, but just few of them are discussed, and the discussions are usually related to a particular implementation of a VR headset brand or visualisation platform. The research of usability and user experience is underdeveloped.

- We observe a huge hype in the quantity of articles, discussing the generative AI, published during the last couple of months. Revealing the real reason behind the enormous interest to the topic is out of the scope of this project but leaves us curious and attracted. No other branch of AI has collected such a mass attention, like gen NLP does. Is it because of awe factor, the technology's usefulness, or just because it manipulates openly with human spoken language, not with numbers and charts, like other in other ML branches?

- The leading analytical agencies, like Gartner and McKinsey report and predict significant acceleration of AI utilisation in business applications. While the number of implementations grows, the proportional risk management factor remains underestimated and mitigation policies undeveloped.

- Human-centred AI (HCAI) and Metaverse are pointed among the four emerging technologies disrupting the next years. The metaverse enables interoperability between digital content and real world's content.

  "The emerging, supporting technologies and trends include (but are not limited to) spatial computing and the spatial web; digital persistence; multi-entity environments; decentralization tech; high-speed, low-latency networking; sensing technologies; and AI applications. " (Nguyen, 2023)

Among data science, ML and AI trends for present and near future Gartner includes these key emerging techniques, in descending order of maturity (*What is AI, 2023*):

- Natural language processing (NLP) as a provider of intuitive forms of communication between humans and systems.

- Knowledge representation - knowledge graphs or semantic networks that facilitate and accelerate the access to and analysis of data networks and graphs. The adoption of knowledge graphs quickly expands over the recent years.

- Agent-based computing - quickly gaining in popularity software agents - persistent, autonomous, goal-oriented programs that act on behalf of users or other programs.

Data Science Holodeck has prioritized the research and development in the above highlighted emerging directions.

Brainstorming

At the stage of ideation and exploration, the project team applied multiple techniques of brainstorming. One if those is the internal enquiry, where the team members elaborated on the scope and priority subjects of research and development. Notes of it are stored in projects artefacts repository (https://innotechspace.dk/holodeck/artifacts/questionnaires/enquiry/).

Prototyping and Experimenting

Three types of prototypes of project products have been created in the process:

- work prototypes, created within the team with a purpose of clarification and validation of ideas;
- test prototypes, used for collecting information and recommendations from potential users;'
- prove of concept prototypes for evaluation and deployment.

The prototype evaluation feedback has been collected by field observation, interviews, and questionnaires. Work hypotheses, design concepts, and code prototypes have been tested experimentally, both in created laboratory conditions and during real practice sessions.

Experiments

One of the early stage software prototypes, developed for validation of the initial ideas can be seen at (Figure 2), as well as at https://innotechspace.dk/holodeck/methodology/prototyping/

It is a Unity application, able to

- input a big data set, used for training ML models
- convert its numeric and text attributes into virtual visuals
- present the visuals in VR, as a 3D scatter plot
- let users interact with the visuals by controller

The users open the VR space, where it is possible to explore the data from a quite different perspective. They are able to move around, to compare the virtual objects, to touch, filter, and move them in the space.

The main purpose of the experiment is estimation of the effect of VR data visualisation on participants' perception and acceptance. It has also been intended as a prove of feasibility. As a hypothesis, a quick orientation and positive acceptance have been expected.

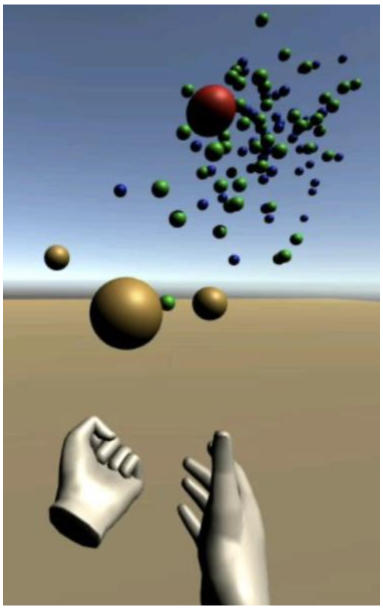

Figure 2 VR Data Visualisation Prototype Source: author

The results of the experiment illustrate various advantages of project's approach to data visualisation, such as faster solving a clustering and recognition tasks, easier conceptualization of the space, much higher motivation and satisfaction of the participants.

Findings

The participants in the experiment with the prototype valued the

- exposition of 3D objects in 3D space
- the multidimensional activation of their perception abilities
- the freedom to interact, control, and change the scene
- the natural way of orientation in the space, multiple perspectives and real perception of scale

The participants contributed significantly to the adjustment of the project tasks specifications with their comments, too. Here are some of the comments we received during the test session:

… "brand new way of learning your data"

… "thinking is slow, seeing is fast"

… "objects draw attention"

… "very intuitive, you can see all immediately"

… "love the look of it"

… "makes sense of otherwise disconnected data"

As a result of the experiments, the team members were able to shape the ideas and clarify the goals of the next research and development stages.

The prototype has been tested and validated as a feasible and proving the concept. It has opened up many directions for future research. It has been used as a basis for further planning and specification of other project developments.

Artificial Intelligence

The development of the project products has taken extensive use of methods, models, and tools from the AI family, including ML, DL, and Gen AI. Here is the distinguishing between them (Figure 3):

Artificial Intelligence

AI (Artificial Intelligence) is a broad field of computer science that encompasses a wide range of techniques and approaches, aiming at creating intelligent machines capable of performing tasks that typically require human intelligence.

Machine Learning

ML (Machine Learning) is a subset of AI that focuses on applying algorithms and developing data models, which allow machines to learn patterns from data, and use these patterns for making predictions and taking decisions. The project has used both supervised ML methods for classification of data and unsupervised for data clustering.

Deep Learning

DL (Deep Learning) is a subfield of ML that involves the use of artificial neural networks with multiple layers (deep neural networks) to process and learn from vast amounts of data. Deep learning has proven to be particularly effective in tasks like image recognition, natural language processing, and other complex pattern recognition problems.

Natural Language Processing

NLP (Natural Language Processing) is a subfield of AI and ML that deals with enabling machines to understand, interpret, and generate human language. NLP utilizes deep learning techniques to process and comprehend textual data.

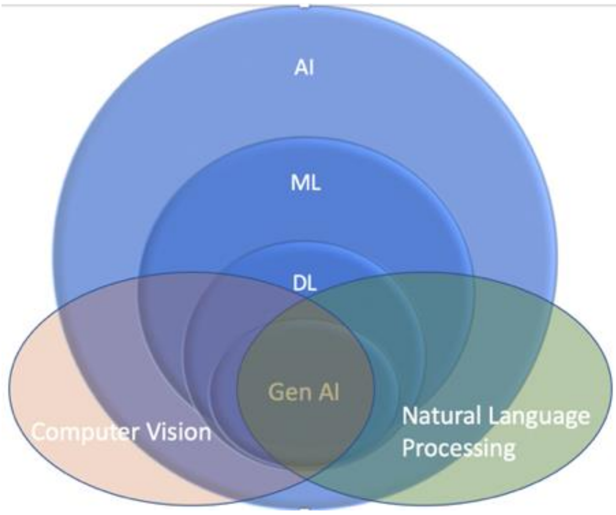

Figure 3 AI Categories Source: author

Generative AI

Gen AI (Generative AI) is a specialized category of AI that focuses on generating new data, such as images, text, audio, or synthetic data, instead of recognizing patterns or making predictions. Generative AI typically involves Transformer architecture and employs techniques from DL, such as generative adversarial networks (GANs), to generate realistic and novel high-quality data that resembles the training data. GAN consists of two neural networks: generator that creates new data and discriminator that evaluates the created data. The generator receives feedback from the discriminator and uses it to improves its outputs. The iterations continue until the generator generates content that is indistinguishable from real data.

Generative AI relies on large models, such as large language models (LLMs) that can classify and generate text, answer questions, and summarize documents.

The projects development workflow incorporates numerous methods and models from the DL, NLP, and Gen AI families. Detailed information about their implementation at technical level can be seen in the software development part of the report (see Product Development Workflow).

Graph Data Science

Knowledge Graphs

KG (Knowledge Graph) is a structured representation of information in a network-like form. It captures data about different entities and their relations in a specific domain (e.g., people, places, and events related to a specific business).

The visual representation of a knowledge graph consists of nodes – the entities, and edges – the relationships between them (see Figure 4).

The project Data Science Holodeck uses knowledge graphs in several roles:

- as a pool of data collected from different sources, but related to one specific business domain or use case;
- as a structure, providing comprehensive knowledge about the domain that is necessary for formulating and answering meaningful and relevant questions or requests for information;
- as a provider of local private knowledge that can be integrated with the external public knowledge, generated by some of the LLMs.

The LLMs, such as GPT, for example, have a vast amount of general knowledge and language understanding. However, they lack access to domain-specific or real-time information. Knowledge graphs can be used in combination with these language models to bridge the gap and enhance local information retrieval.

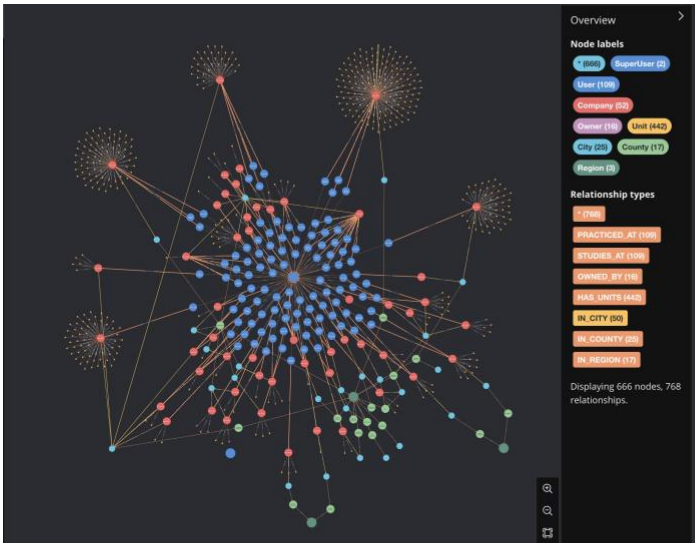

Figure 4 Knowledge Graph in Neo4j *Source: author*

Once the knowledge graph of domain-specific data is created, the software application can invoke LLM as a better facilitator of the dialog between the user and this KG. For example, the LLM can be used to understand the intend of the user or the context of the query, as well as to identify key pieces of information inside the knowledge graph that might be essential for the response

In another scenario, the LLM can leverage the knowledge graph's information and this way to become able to retrieve relevant local information.

By combining the strength of LLM and its generative capabilities with the structure and the relevance of the domain-specific information stored in the knowledge graph, the process of retrieving local information becomes more efficient, accurate, and up to date.

Graph Algorithms

A valuable advantage of using graph structures of data instead of the traditional SQL and NoSQL databases is the possibility to benefit from applying algorithms, known from the mathematical graph theory.

The project Data Science Holodeck implements graph algorithms for operations with and analysis of the knowledge graphs mentioned above. Some such algorithms are named on Figure 5 below.

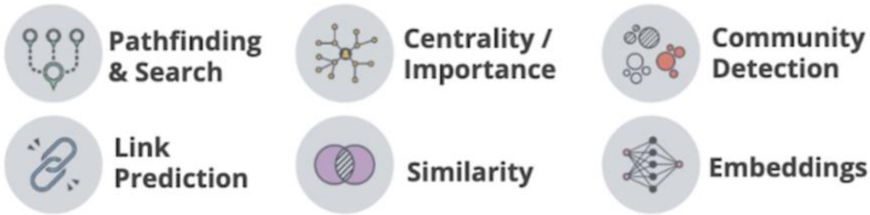

Figure 5 Categories of Graph Algorithms *Source: Neo4j*

HCI Evaluation

HCI evaluation is an important part in project's process and usability is a major feature of project's products, therefore a range of methods for usability evaluation have been implemented. Factors that have impact on the users' performance and quality of experience have been explored in other experiments. The focus has been placed on revealing the success factors, preferences, and challenges experienced by the users interacting with the VR system. This includes human factors, as well as VR factors, such as quantity and density of visuals; volume, scaling, and organisation of the space; velocity and dynamics on the scene; individualisation and means of help.

Further, usability research on the level of cognitive support has been performed. It involves assessment of intuitiveness of operating, efficiency, and satisfaction of participants.

HCI evaluation has been a continuous iterative process, going in rounds of testing and improvement of the stage and the interactions design. One example of subjective feedback collection is stored in projects repository (https://innotechspace.dk/holodeck/artifacts/ questionnaires/feedback/).

As a result of analysis of collected data from all experiments, interviews, and questionnaires, a list of VR UX Design Recommendations has been published in the Appendix VR UX Design Recommendations. A VR demo implementation has also applied the recommendations (see https://innotechspace.dk/holodeck/demo-cases/case-1/).

Use of Technology

Data Science Holodeck is a technology-based project. The research and development process require setting and implementation of numerous computer technologies, both hardware equipment and software instruments.

Software Framework

The relevant theoretical concepts, the accepted design solutions, the significant outcomes, and their illustrating implementations have been validated in practice and integrated in a configurable software system, built of modules.

The system is programmed in Python and includes a framework of components necessary for creating AI applications with VR visualisation functionalities. Similar to building with Lego blocks, the Holodeck system components integrate in a configuration that is best suited for a specific business case.

The workflow of a particular configuration can vary depending on the task but starts with collecting and loading domain-specific data from variety of sources into a knowledge graph, from where gets analysed and transformed into relevant structures, following appropriate AI models. Finally, the data is visualized in 3D and VR, and used for solving business tasks.

Detailed description of a generic AI-VR workflow and its software library prerequisites is provided in the Appendix Product Development Workflow.

Hardware Platform

The system can run, in different versions, as a web-based application in a web browser or as a Unity application on Android-based platform. In both versions it requires a computer with high processing power and eventually, larger storage space for the data.

The VR visualisation requires also use of VR headsets, VR controllers, and connectors. Depending on the configuration, Internet connection may or may not be needed.

Data

The system runs with the data, provided or pointed by the user. It can operate with and integrate large amount of data sources, types and formats, such as PDF, DOC, TXT, CSV, JSON, HTML. It 'understands' numeric, text, image, video, and audio documents.

Some of the typical uses of the collection are extracting of information, creating summaries, and answering questions in a natural language. The system can work in Danish but produces better results in English language.

PROJECT OUTCOMES

The project Data Science Holodeck provides research and development of innovative methods and tools that can facilitate the process of collecting, exploring and understanding complex data. It investigates and integrates advanced artificial intelligence methods with modern virtual visualisation techniques in support of solving the complexity. The main outcomes of the project work include:

- Product Development Workflow: A methodology for exploration and visualisation of data and operations based on the recent achievements in data science, artificial intelligence, and virtual reality subject areas. The methodology considers a variety of business cases of data processing and analytics and provides generic approach, methods, and instruments in support of their solutions (Product Development Workflow).
- Software Framework: A library of software instruments for implementation of the methodology that can be used for ingestion, exploration, and visualisation of wide range of sources within a certain domain of human activities. It enables configuring of different tasks and solutions related to data analytics and visualisation that can be used as illustration and dissemination of AI and VR knowledge and implementations (Product Development Workflow).

- VR Demo Primers: A set of programming applications of demo cases, explaining the concepts and the qualities of project's achievements. The primers serve as a prove of concept but can also be used as guides and foundation of further development, adaptation, and dissemination of AI and VR support (VR Demo Primers).

- VR UX Design Recommendations: A list of recommendations for design, development, and use of 3D and VR user interfaces, as well as for creating positive user experience, while solving data related tasks in business and education. The recommendations are generated by own experiments and validation, and contribute to the collective experience in the area (VR UX Design Recommendations).

- Glossary of Terms: Plain explanation of technical terms and advanced concepts, carrying ground meaning in the vocabulary of Data Science Holodeck. The explanation addresses users with non-technical background or limited knowledge and experience in AI and VR (Glossary of Terms).

The project products would have a wide effect in support of:

- democratisation and popularisation of the emerging information technologies AI and VR to SMEs (micro, small, and medium size business units) and individuals, reducing significantly the onboarding time or the time to taking decisions related to the implementation of these technologies;
- development of AI-related content and advanced methods of teaching and training in higher education for profession of high demand;
- further research and innovations in the area.

The expected implication of the results on business would be related to easier processing and better understanding of large or complex data sources, optimisation of decision-making process, significant reduction of both human and computer time used for working on data and document related tasks, faster implementation of AI technology and applications.

The immediate benefit in education would be the use of innovative and smart teaching methods, tools, and materials, in achieving higher learning goals in response of the technological challenges and the expectations of business and society.

FURTHER DEVELOPMENT

Data Science Holodeck works as a pilot exploration and implementation stage of the two powerful advanced technologies – Artificial Intelligence and Metaverse, which right now are transforming the whole IT space and its traditional integration with the business and society.

The project achievements form a solid base for informed beginning of accommodation of the innovations in certain academic and small business practices. They possess the potential to configure and adapt to various business needs and usage scenarios that could be the next right step towards broader distribution and larger effect.

Data Science Holodeck 2.0 would be a great opportunity for proper dissemination and incorporation of the results in real-live implementations.

The project is a 'living' structure, where further research and development are applied on a permanent basis, in accordance with the higher dynamics of innovations in the underlying science and technology areas.

CONCLUSION

The idea of integrating AI and VR for support of data operations and cognitive activities in business and education is still quite new and unexplored at local and global scale.

Both the process and the products of the project contribute significantly to the individual and collective expertise in research, development, and implementation of these rapidly expanding technologies in business and education.

Being multidisciplinary by foundation, the project objectives address not only IT, but business areas, as well. It can be regarded as a bridge between the two directions and used as a mean of knowledge transfer.

The project outcomes create an opportunity for increasing the overall awareness and speeding up the distribution of AI and VR related knowledge and implementations across the school and enterprises.

APPENDICES

- Product Development Workflow
- VR Demo Primers
- VR UX Design Recommendations
- Glossary of Terms
- Reference to Resources

ARTEFACTS

- Questionnaires
- Open Session
- Gallery