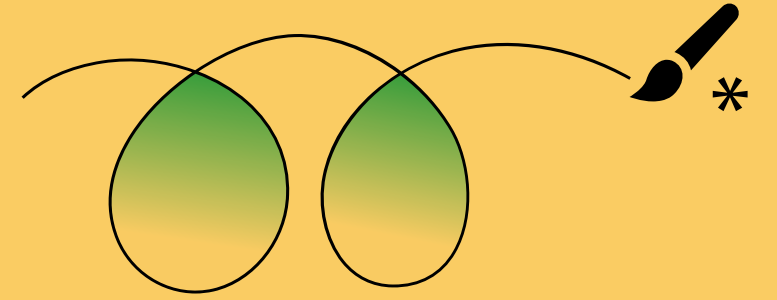
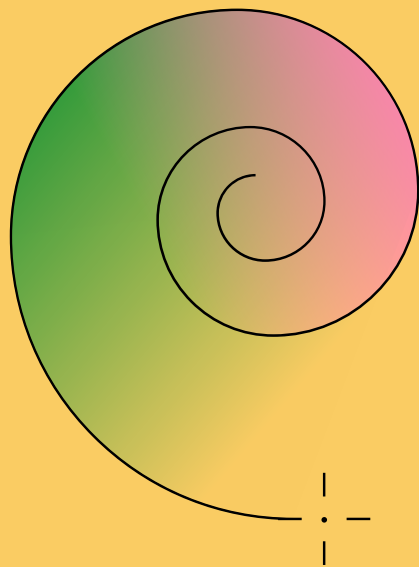


UNICAM



Technology for Big Data



Memory Store

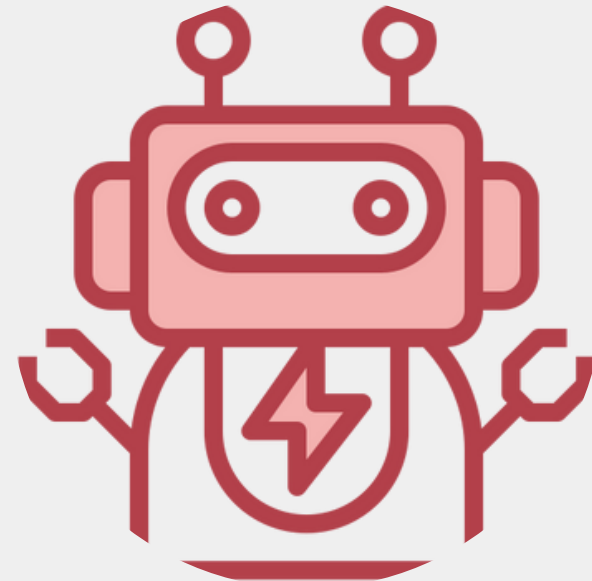
Instructor: Prof.Massimo Callisto De Donato
Student: Van Hieu Ho



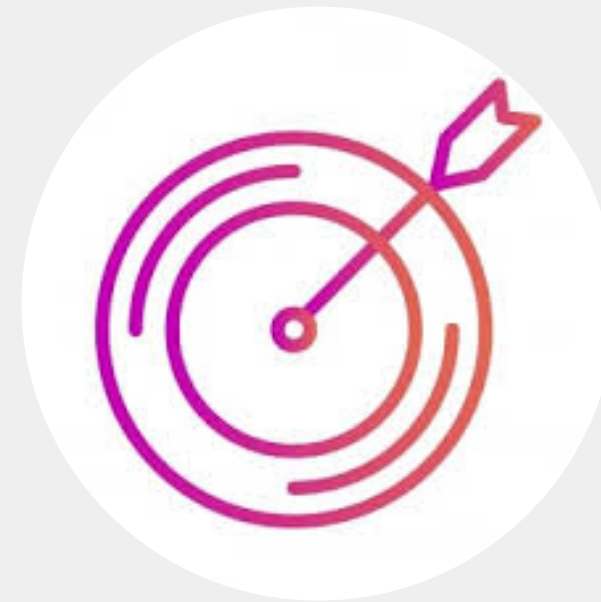
Outline



Introduction



**Technology
and Function**



**Further
Development**



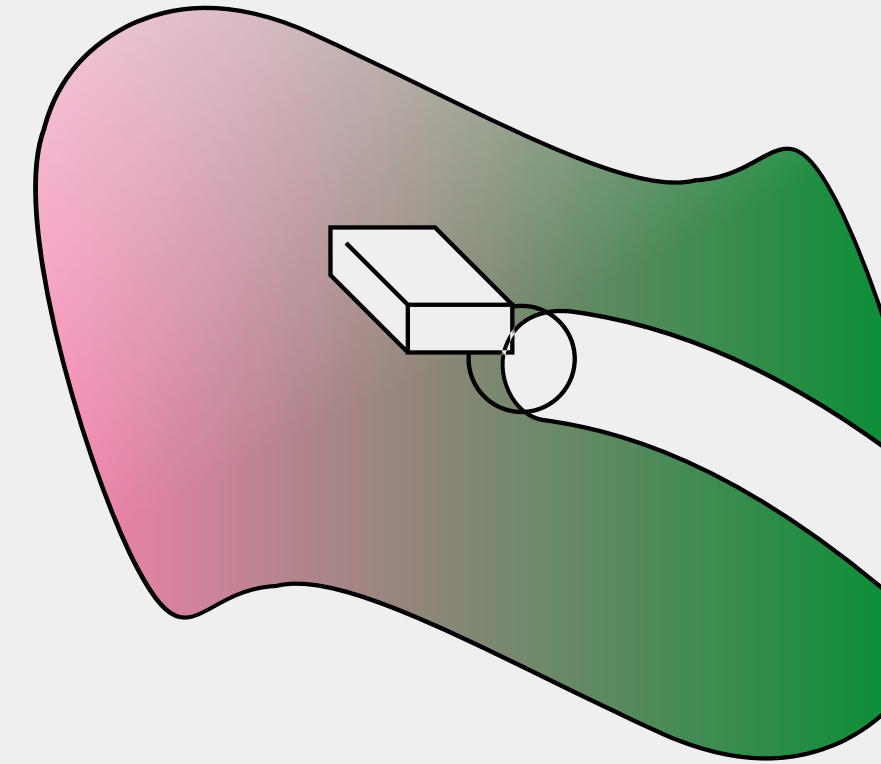
Testing

Introduction

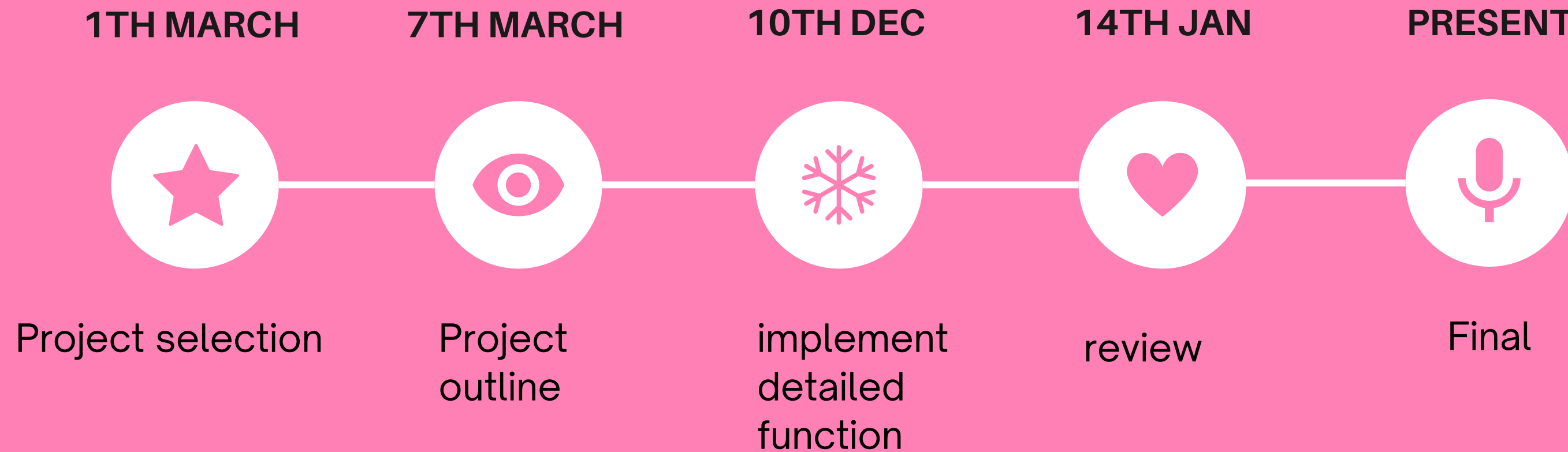
Overview of the project and its purpose

Main objectives:

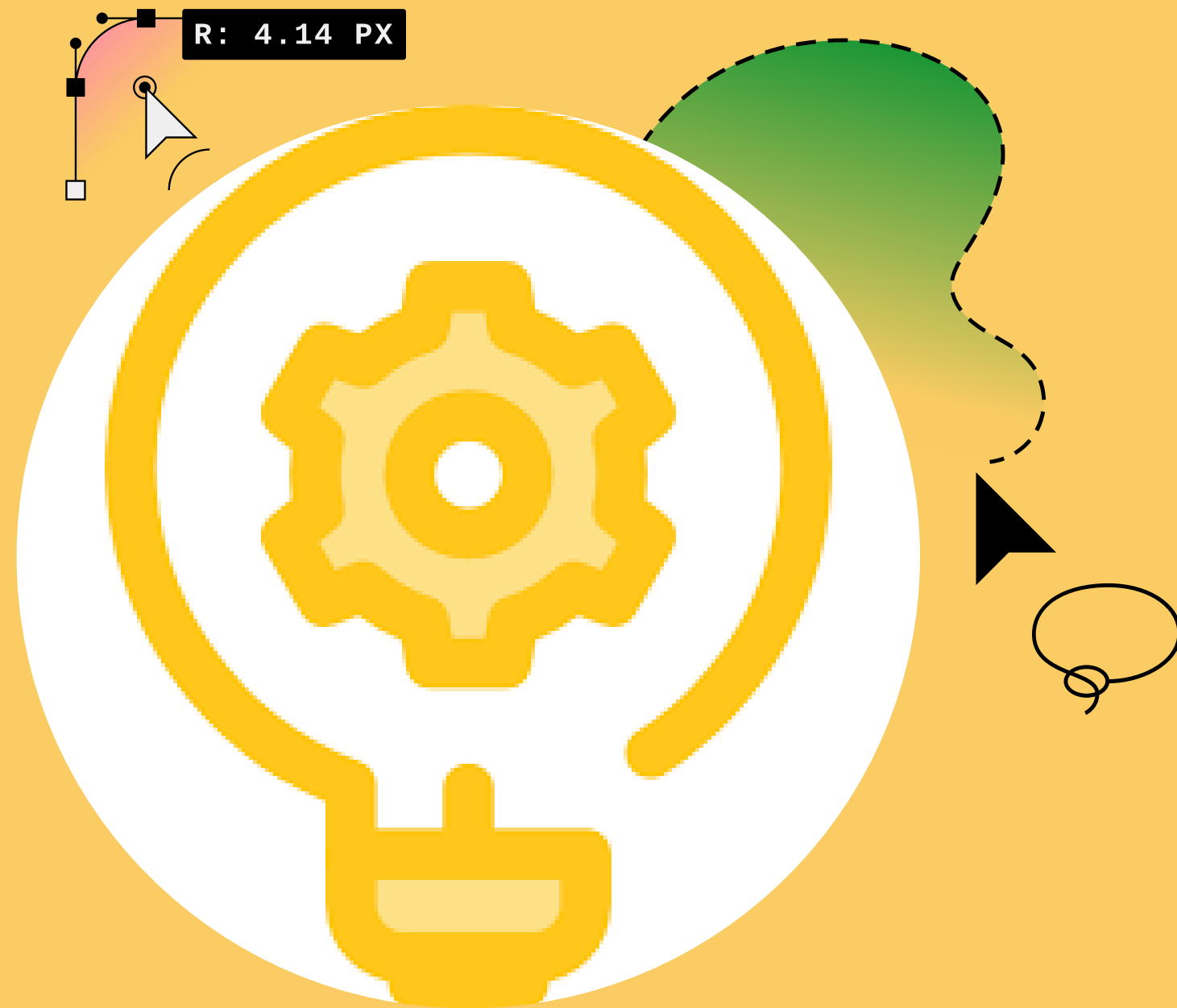
- Collect data from CEUR.
- Store it in MongoDB.
- Visualize the data.
- Implement advanced search using LLM-Vector Search.



Development Phase



Technology and Function



2.Technology

2.Function, Explanation

2. Technology Used

1. Data Collection & Storage

- Crawl data from CEUR and store it in MongoDB.

ceur-ws.org
ISSN 1613-0073

CEUR Workshop Proceedings (CEUR-WS.org) is a free, diamond open-access publication service at Sun SITE Central Europe operated under the umbrella of RWTH Aachen University and listed in the BioRxiv OA dataset. CEUR-WS.org is a recognized ISSN publication series, ISSN 1613-0073 (json). CEUR-WS.org is hosted at <http://SunSITE.Informatik.RWTH-Aachen.DE/Publications/CEUR-WS/>. This service is provided by the CEUR-WS.org Team. See end of the page for contact details and Impressum.

[submit](#) | [GenAI Policy](#) | [AixIA Series](#) | [IAOA Series](#) | [Blog](#) | [Long-term archive](#) |

op Proceedings (CEUR-WS.org)

Proceedings for Computer Science Workshops

formance to the legal Disclaimer of Sun SITE Central Europe (CEUR) and the legal Disclaimer of Technical University of Aachen (RWTH), the copyright for the workshop proceedings as a compilation, is with the respective proceedings editors. The copyright for the individual *items* (subsuming any type of computer-represented files containing articles, software demos, videos, etc.) within the proceedings volume remains with the authors/owners. The open-access license for a volume is specified in the index file of the respective volume. This license applies by default to all components in the volume. **Re-publication** of a CEUR item inside a proceedings volume requires permission by the copyright owners, i.e. either the respective proceedings editors, or the authors of the respective item in that volume, or both. Mirror sites are prohibited. The label 'CEUR Workshop Proceedings' and the CEUR-WS logo are owned by the members of the CEUR-WS Team, represented by its editor-in-chief. CEUR-WS.org provides its services as a free of charge service. CEUR-WS.org is not run by an organization but by volunteers from different universities, who realize the service in their spare time.

* timeline. We are grateful for donations of scripts that ease our tasks, for example scripts that detect errors in index files.

ow our [instructions on how to submit your proceedings volume](#).

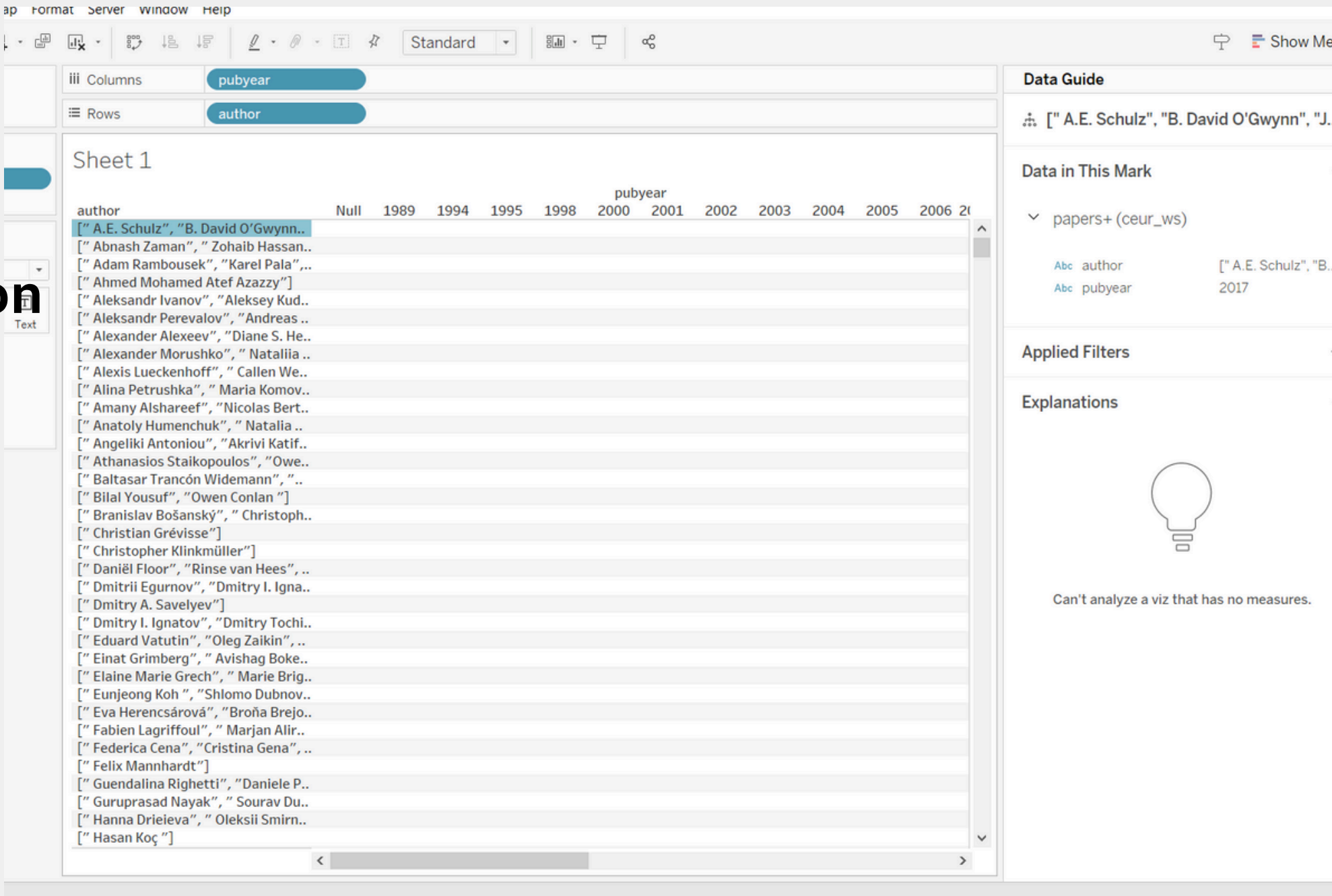
g war in Ukraine, CEUR-WS suspends until further notice submissions from Russian and Belarusian institutions.
mplate for the CEURART style at Vol-XXX was updated to match more closely the LaTeX originals.
s a new policy about using Generative AI in papers. This policy mandates that authors: i) declare the use of GenAI tools in the paper, ii) review and edit any AI-generated content. [Read more](#).

ence for Agriculture in Africa 2024.
ings of the Workshop on Data Science for Agriculture in Africa (DAAfrica 2024), Bejaia, Algeria, November 23, 2024.
y: Paulin Melatagia Yonta, Mathieu Roche
d by: Mathieu Roche
d on CEUR-WS: 17-Mar-2025
: <http://ceur-ws.org/Vol-3944/>
nbn:de:0074-3944-X

2. Technology Used

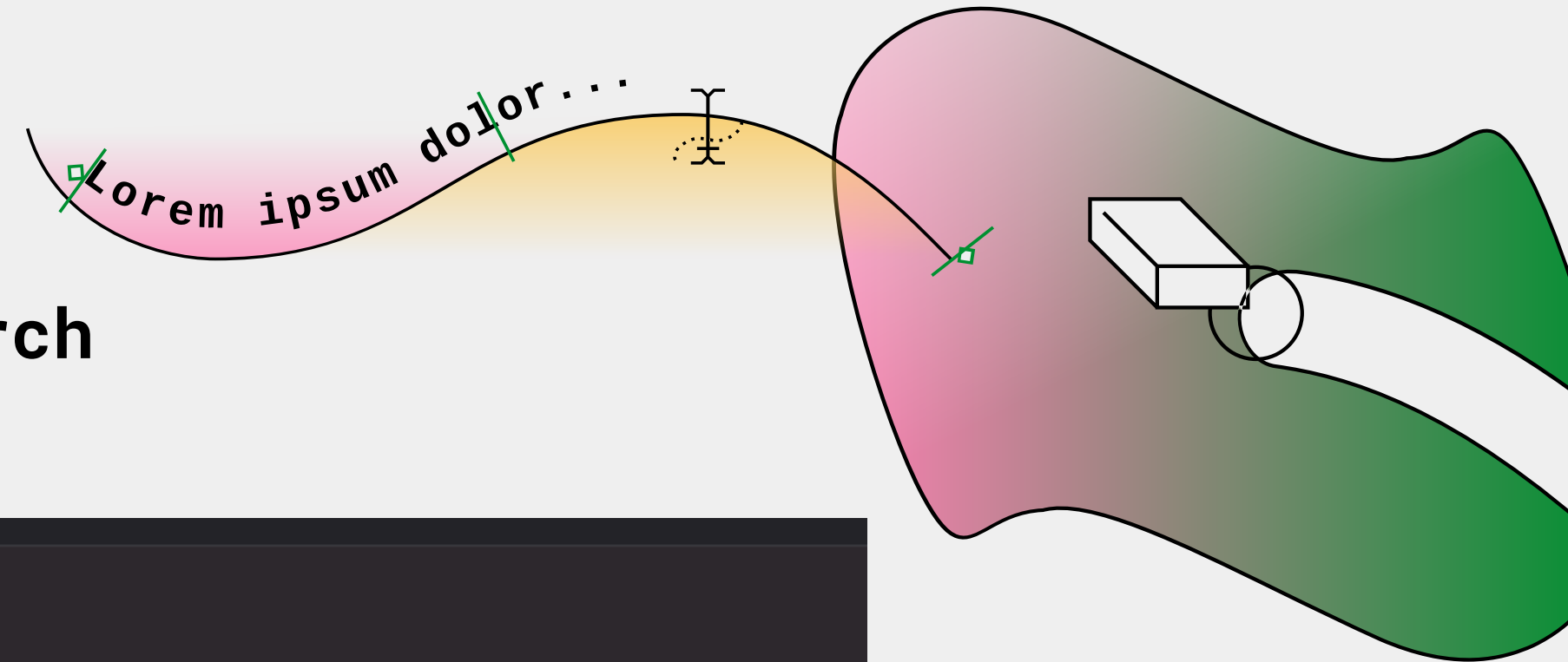
1.Data Processing & Transformation

Convert and preprocess data for visualization.



2. Technology Used

1. Query Data based on llm-vector-search



```
[3]
.. Pinged your deployment. You successfully connected to MongoDB!

> v
# Fetch the first document from the collection
db = client.get_database('ceur_ws')
papers_collection = db.papers

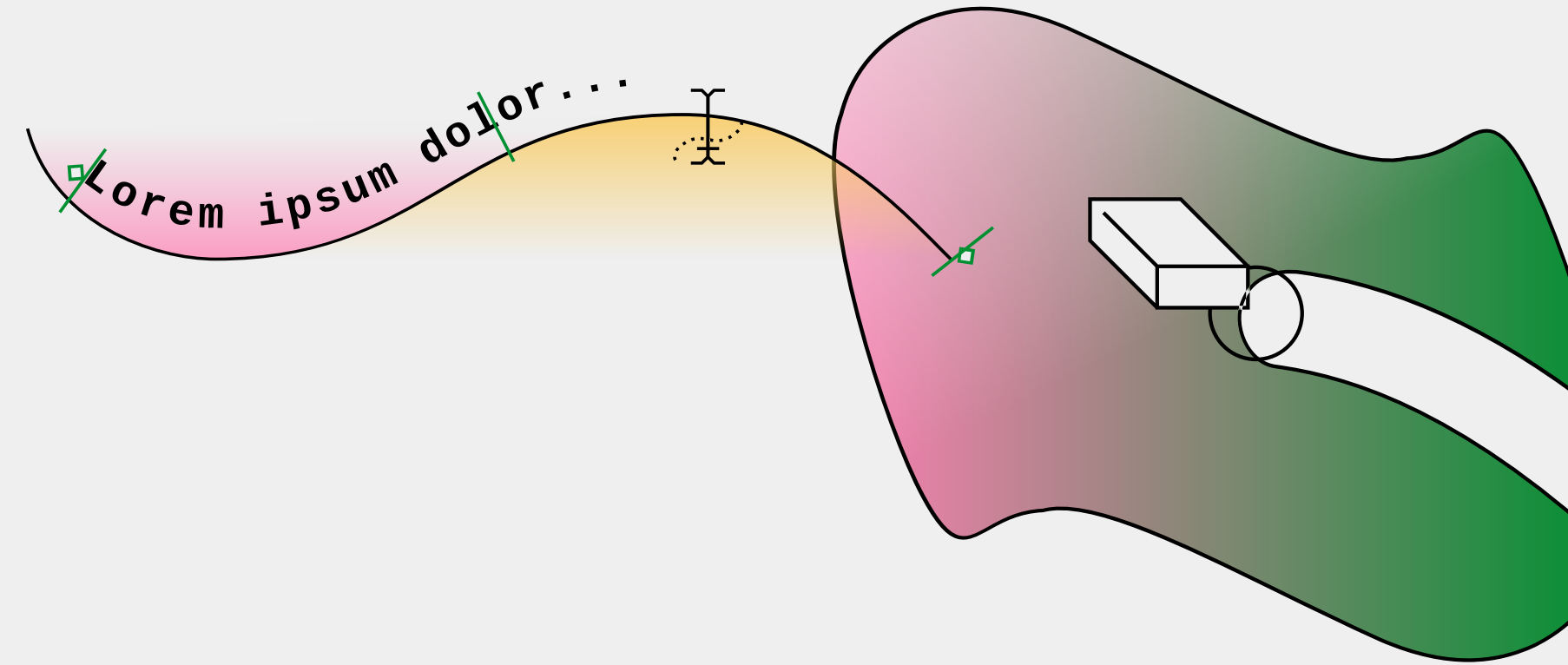
title = "Empowering Supply Chains Resilience: LLMs-Powered BN for Proactive Supply Chain Risk Identification"
document = papers_collection.find_one({'title': title})
print(document)

[4]
.. {'_id': ObjectId('67d7dec76dba08ca73f92681'), 'url': 'https://ceur-ws.org/Vol-3707/D2R224_paper_2.pdf', 'title': 'Empowering Supply C
```


2. Technology Used

1. Data Processing & Transformation

- Implementing vector-based retrieval for semantic search.
- How LLM and embeddings improve search accuracy.

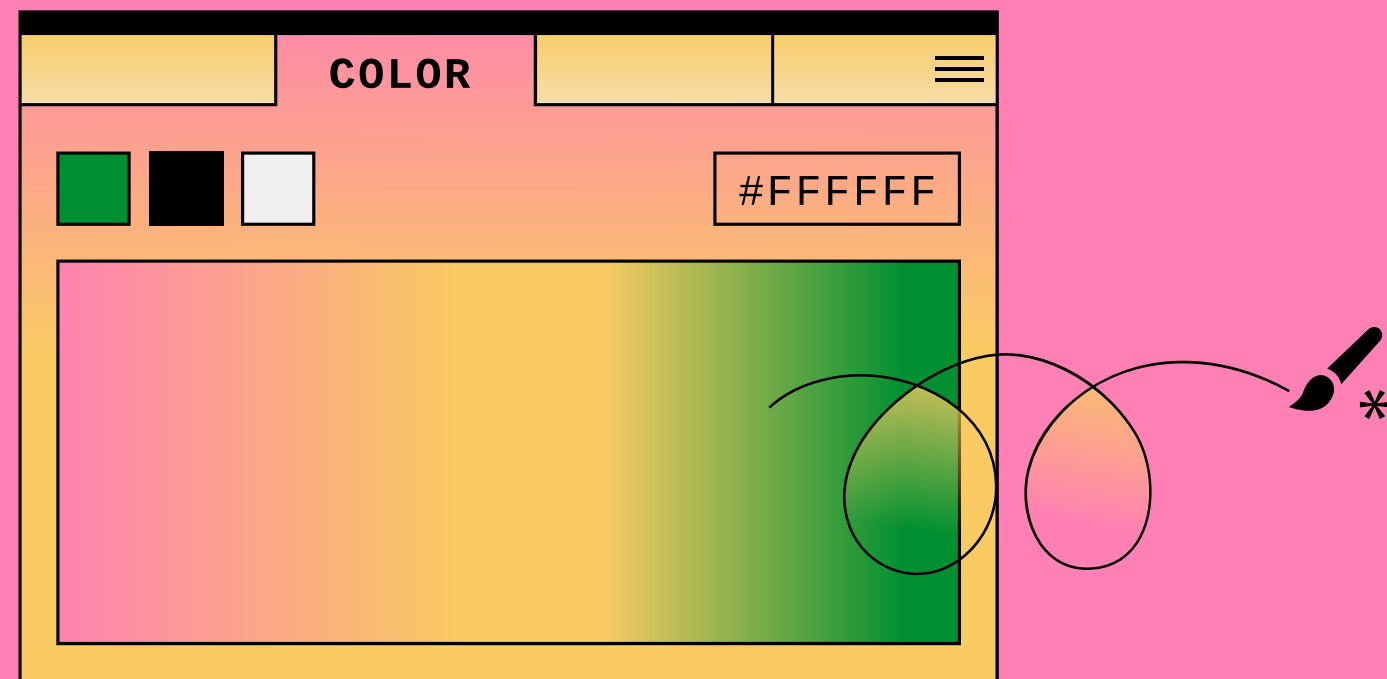


Further Development

1. Fully system with orchestration function

2. Advanced Search with LLM-Vector Search:

- Implementing intelligent search with Vector Search.



Testing and Question

