



**Hochschule  
Bonn-Rhein-Sieg**  
*University of Applied Sciences*

Fachbereich Informatik  
*Department of Computer Science*

# Project Plan

Visual Computing and Games Technology (MSc)

## Tell me mAI story

by

**Vanessa S.**

**Sven L.**

**Hoan V.**

2023-12-07

# 1 Work Packages

## 1.1 Functional Area - Processing and Storage of Children's Drawings

This section covers the topic of ingesting new previously unseen children's drawings, i.e. when a child has created a drawing and first presents it to our AI.

### 1.1.1 Work Package - Basic dialogs

Upon starting, the tool welcomes children. In one of the available basic dialogs, it explains some of the concepts. In another basic dialog, it presents a catalog of supported animals, with examples. In the drawing upload dialog, it accepts an image file, in a format like JPEG or PNG. One of these scenarios could happen:

- If there is no detectable object, tell the user to draw an animal chosen from the catalog. (Note that the term object here refers to any recognized object from the drawing, it could be anything from a shape to an animal that is from the catalog or not.)
- If there are detected objects, based on the classification score for each object, verify with the user if the classification is correct, by presenting the detected animal region and the predicted animal name, any by asking a question.

The tool provides a friendly farewell when the interaction session concludes.

### 1.1.2 Work Package - Curation of the training dataset of drawings

First batch of dataset is collected from an elementary school. To prepare the dataset for training, some preprocessing steps are required:

- Resize the image to a fixed size.
- Determine and annotate bounding box coordinates and labels of animals in a file (for example in CSV format) for supervised learning.
- Split the dataset into default training, validation, test set. Note that later we might make use of random splits or even cross validation.

For a wider range of recognition ability and for improved accuracy, we could make use of public datasets. Some available datasets are:

- [Google Quick, Draw! dataset](#) (hand drawn doodles)
- [TU-Berlin Human Sketch Recognition](#)

We also downloaded some examples from the official media library at the university using the student license.

### 1.1.3 Work Package - Experiments with preprocessing of drawings

In this work package we refine the preprocessing of drawings.

### 1.1.4 Work Package - Preprocessing pipeline of drawings

In this work package, we create the preprocessing pipeline of drawings that is to be part of the tool.

## 1.2 Functional Area - Recognition of animals

This section covers the solution to the problem of recognizing animals in children's drawings, that is recognizing that an animal is present and classifying the animal.

### 1.2.1 Work Package - Experiments with a first CNN for recognizing animals

Implement a first CNN for recognizing animals in terms of a binary classification (animal present vs. animal not present) and evaluate its performance. Frameworks could be tested are: PyTorch, TensorFlow, Keras, etc.

### 1.2.2 Work Package - CNN for recognizing animals in terms of a multi-class classification

Implement a CNN for recognizing animals in terms of a multi-class classification, i.e. detecting an animal and classifying it, and evaluate its performance.

### 1.2.3 Work Package - CNN for recognizing multiple animals in a single drawing

Implement an advanced CNN model (e.g. Faster R-CNN, Mask R-CNN) for recognizing multiple animals in a single drawing, localize the animal with bounding boxes and classify each one of them. Evaluate its performance.

## 1.3 Functional Area - Advanced recognition of animals (optional)

This section is optional. We want to look at alternative advanced models for recognizing animals in children's drawings, in particular we want to consider recent state-of-the-art models and concepts that have been developed by researchers and companies in the years 2020-2023.

### 1.3.1 Work Package - Theoretical consideration of alternative models for animal recognition

- ConvNet: [arXiv link](#)
- CoTNet: [arXiv link](#)
- VAN: [arXiv link](#)
- MAE: [arXiv link](#)
- ViT: [arXiv link](#)
- Faster R-CNN [arXiv link](#)
- YOLO (lightweight, real time recognition but might be not as accuracy as Faster RCNN) [arXiv link](#)
- EfficientDet (lightweight, suitable for small size dataset) [arXiv link](#)

## **1.4 Functional Area - Annotating animals with digits to denote their importance**

This section covers the topic of annotating animals with digits to provide the model a clue about their importance in the story. An animal annotated with a 9 would be a main character, while an animal annotated with a 1 would just appear somewhere in the background of the story.

### **1.4.1 Work Package - Model for recognizing handwritten digits in terms of a multi-class classification**

Implement a model for recognizing handwritten digits in terms of a multi-class classification, i.e. detecting a digit and classifying it, and evaluate its performance.

### **1.4.2 Work Package - Advanced model for recognizing multiple handwritten digits in a single drawing**

Implement an advanced model for recognizing multiple handwritten digits in a single drawing, localize the digits with bounding boxes and classify each one of them. Evaluate its performance.

### **1.4.3 Work Package - Associate recognized digits with recognized animals**

Using a measure of shortest spatial distance in an image, i.e. between each digit and animal, or some other applicable measure, possibly using a clustering method, each recognized digit is associated with one of the recognized animals, in such a way that every animal is associated with exactly one digit. If there are not enough digits, the default digit 5 is assumed for all animals that have no digits, and a warning is written to the log. If there are too many digits, digits with the greatest spatial distances to any animals are ignored, as many as applicable, and a warning is written to the log. In normal cases, each animal is associated with at most one digit, and each digit is associated with exactly one animal. It is a 1-to-1 relationship between digits and animals. More concretely, 1 animals to 0..1 digits. Only in the special case where there are more digits than animals in the drawing, a digit may be associated with 0 animals.

## **1.5 Functional Area - Annotating animals with markers to vaguely denote their relations**

Markers are used to indicate which characters should meet with which other characters during the story. Note that, while more elaborate functionality could be conceived, for example a formation of fixed groups of animals, or a notion of typed relations between animals, such as friends, family, colleagues, in the case of this project, the markers just stand for a vague association. This is to give the AI a lot of room for interpretation, or better said, more freedom for the generation of the story.

### **1.5.1 Work Package - Evaluate different marker types and choose the marker type to use**

Different marker types need to be considered. A particular applicable marker type is to be chosen. A respective evaluation of markers, on which the decision can be based, may

be kept simple and brief. For the marker type, the most probably choice is colored boxes, e.g. empty post-it notes, because such are usually available in many households, and if not, can be produced easily by the children or the parents. Animals with markers of the same color would be vaguely associated.

#### **1.5.2 Work Package - Model for recognizing markers in terms of a multi-class classification**

Implement a model for recognizing markers in terms of a multi-class classification, i.e. detecting a marker and classifying it, and evaluate its performance.

#### **1.5.3 Work Package - Advanced model for recognizing markers in a single drawing**

Implement an advanced model for recognizing multiple markers in a single drawing, localize the markers with bounding boxes and classify each one of them. Evaluate its performance.

#### **1.5.4 Work Package - Associate recognized markers with recognized animals**

Each animal is associated with zero or more markers, while each marker is associated with exactly one animal. It is a 1-to-n relationship between animals and markers. More concretely 1 animals to 0..n markers.

#### **1.5.5 Work Package - Determine groups of animals with respect to the recognized and associated markers**

Using a measure of shortest spatial distance in an image, i.e. between each marker and animal, or some other applicable measure, possibly using a clustering method, each recognized marker is associated with one of the recognized animals, in such a way that every animal is associated with zero, one or more markers. Each animal is associated with zero or more markers, while each marker is associated with exactly one animal. It is a 1-to-n relationship between animals and markers. More concretely 1 animals to 0..n markers.

### **1.6 Functional Area - Annotating animals with activity cards to denote their actions**

Pre-defined activity cards can be placed onto the picture to indicate which characters should engage in which activities (e.g. eating, drinking, jumping, going for a walk) during the story. (The activity cards are considered a potential feature, still to be prioritized with respect to other features and details.)

#### **1.6.1 Work Package - Sketch and design various activity cards**

Ideate and prototype activity cards either physically or digitally.

#### **1.6.2 Work Package - Experiment with various styles in terms of art**

Experiment with different art styles. The art style should support cases where only a grayscale or black and white printer is available to the parents.

### **1.6.3 Work Package - Create the activity cards digitally**

The activity cards need to become digitalized in terms of print media.

### **1.6.4 Work Package - Define and make first use of a process for producing physical activity cards**

It must be possible for parents to print and cut out the activity cards.

### **1.6.5 Work Package - Model for recognizing activity cards in terms of a multi-class classification**

Implement a model for recognizing activity cards in terms of a multi-class classification, i.e. detecting an activity card and classifying it, and evaluate its performance.

### **1.6.6 Work Package - Advanced model for recognizing activity cards in a single drawing**

Implement an advanced model for recognizing multiple activity cards in a single drawing, localize the activity cards with bounding boxes and classify each one of them. Evaluate its performance.

### **1.6.7 Work Package - Associate recognized activity cards with recognized animals**

Using a measure of shortest spatial distance in an image, i.e. between each activity card and animal, or some other applicable measure, possibly using a clustering method, each recognized activity card is associated with one of the recognized animals, in such a way that every animal is associated with zero, one or more activity cards. Each animal is associated with zero or more activity cards, while each activity card is associated with exactly one animal. It is a 1-to-n relationship between animals and activity cards. More concretely 1 animals to 0..n activity cards.

## **1.7 Functional Area - Generating a children's story based on a children's drawing in text form**

Based on the information gained by analyzing the drawing, a children's story is generated and then told.

### **1.7.1 Work Package - Create internal data structure to represent an annotated drawing**

Gained information, such as animal types, spatial locations of animals in the drawing, importance of each animal obtained via digits, associations between animals obtained via markers, activities of animals obtained via activity cards, must be written to a data structure. An internal data structure is facilitated that is then used as input to the story telling module.

### **1.7.2 Work Package - Experiment with free to integrate LLMs**

Determine how to use and how to integrate LLMs, and which LLM to use.

### **1.7.3 Work Package - Engineer prompts to make the LLM tell good stories**

For the LLM to tell a good children's story based on the input, we need to apply instruction tuning and prompt engineering.

## **1.8 Functional Area - Tell children's story using text-to-speech**

To actually tell the story, we need to apply a text-to-speech module and integrate it into the tool.

### **1.8.1 Work Package - Apply text-to-speech**

We prepared this work package in another project. We need to integrate and apply the solution.

## **1.9 Functional Area - Scientific research**

Continuous research for different aspects of the project for different written assignments. Possible topics, associated with certain functional areas or technical approaches considered during the project, may include:

- Deep learning approaches for analyzing children's drawings
- Generative AI approaches for creative storytelling with a focus on children's stories
- User interfaces for interaction between children and AI systems

### **1.9.1 Work Package - Placeholder 1 for scientific research**

### **1.9.2 Work Package - Placeholder 2 for scientific research**

### **1.9.3 Work Package - Placeholder 2 for scientific research**

## **1.10 Functional Area - Project evaluation**

Finally, the created product is evaluated.

### **1.10.1 Work Package - Preparation of evaluation method**

This includes the preparation of a consent form for the parents, the acquisition of child testers and scheduling of the evaluation.

### **1.10.2 Work Package - Evaluation of the product**

Evaluation of the project including drawing of animals, recognizing by the system, annotation with different markers and the prompt of the final story. The aim is for a test group of 10 to 15 or more children.

### **1.10.3 Work Package - Analysis of evaluation feedback**

Collecting, analyzing and processing the evaluation data.