

# Keeping Waterways Clean with ML

Project with Singapore's Public Utility Board

Paola Bianchi, Silvan Melchior | SDS 2024

# About us



Dr. Paola Bianchi

Lead Data Scientist

@ maria-paola.bianchi@zuehlke.com

 <https://www.linkedin.com/in/paolabianchi>



Silvan Melchior

Lead Data Scientist

@ silvan.melchior@zuehlke.com

 <https://www.linkedin.com/in/silvan-melchior>



# About PUB

The Singapore's Public Utility Board is responsible for

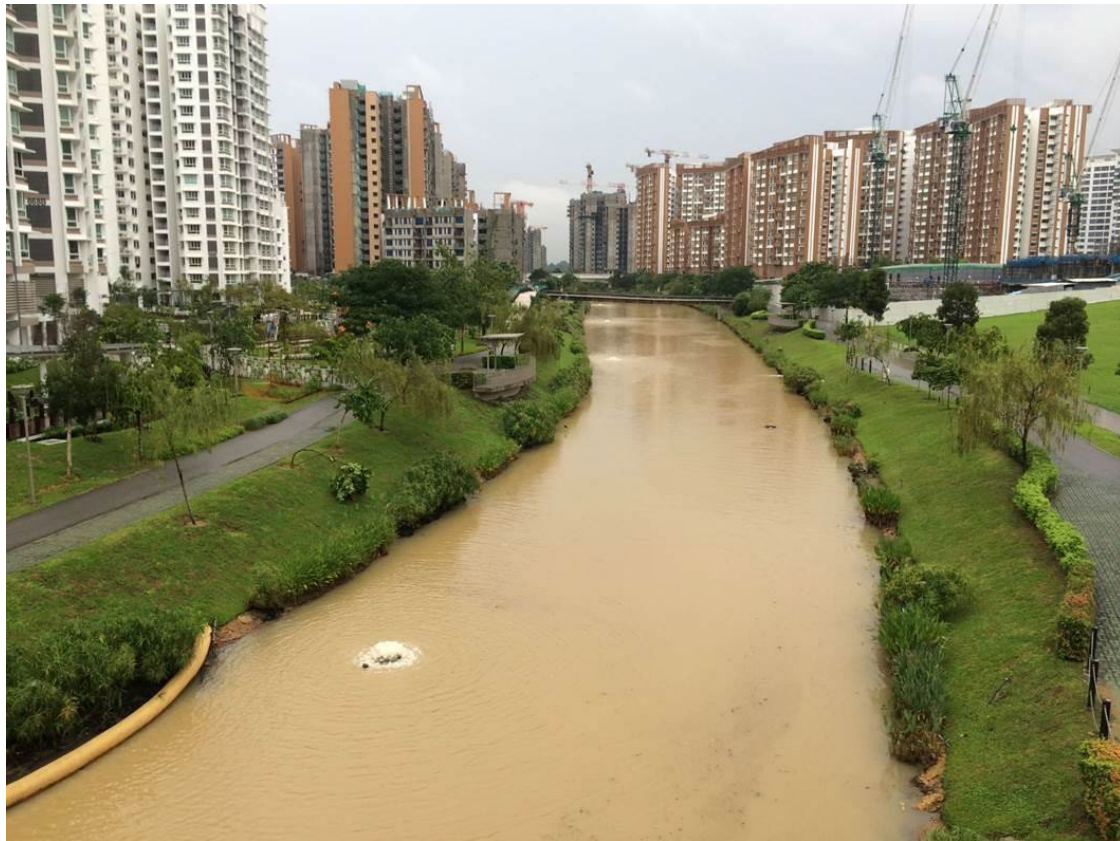
- Water supply
- Water catchment
- Used water





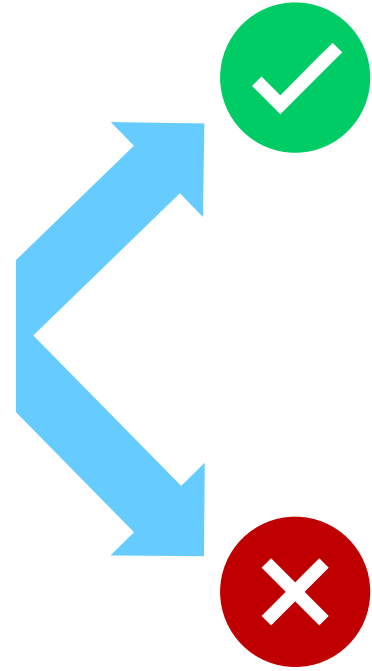
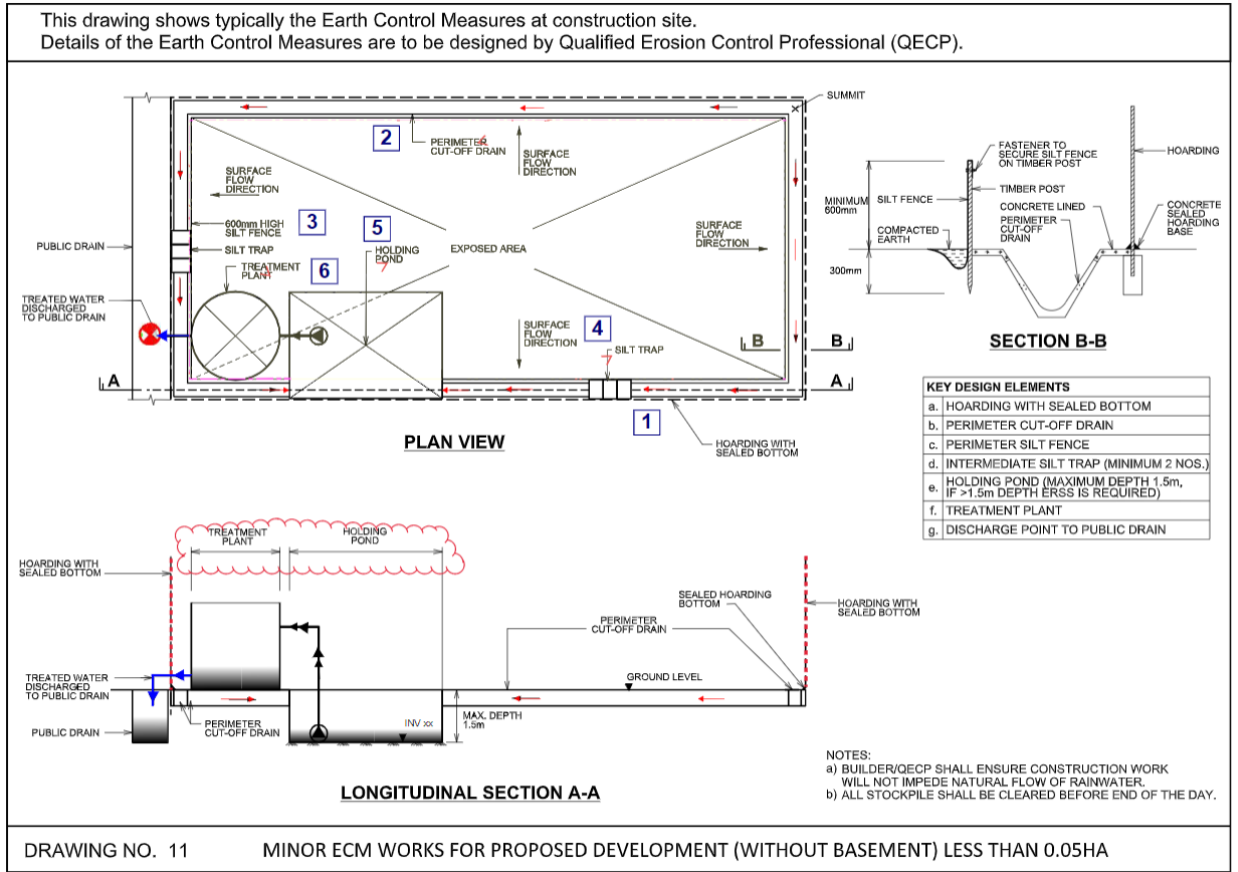
# PUB's Challenge

# Earth Control Measures



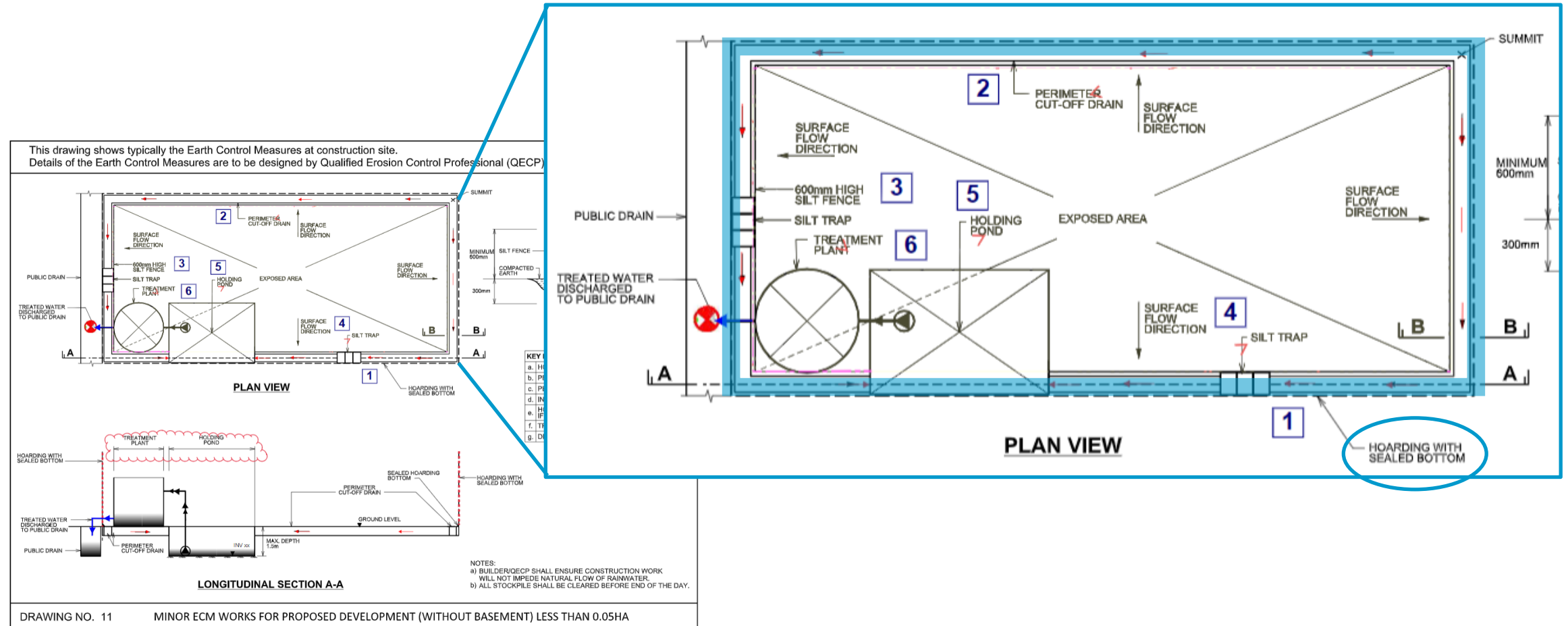
# ECM Plan checking

## Expected submission



# Example of rule checking

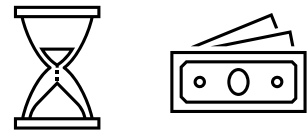
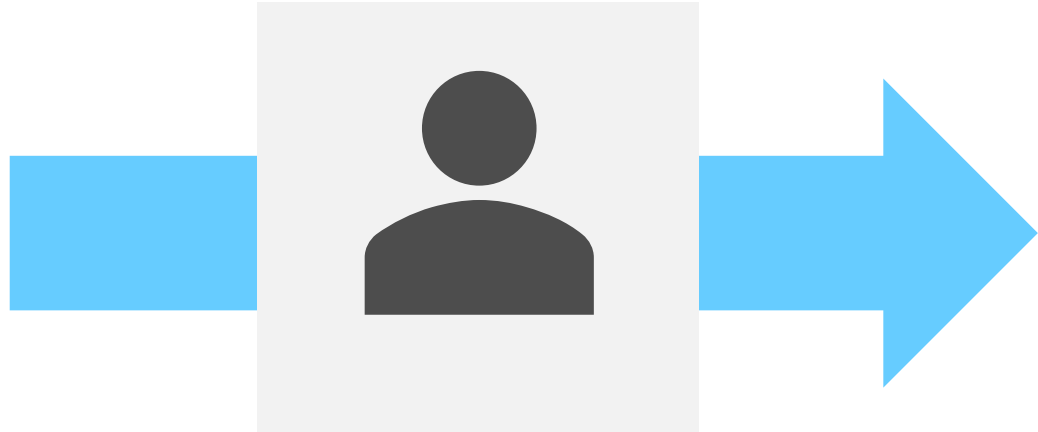
Rule: hoarding forms a closed loop and is sealed at the bottom



# ECM submission check process

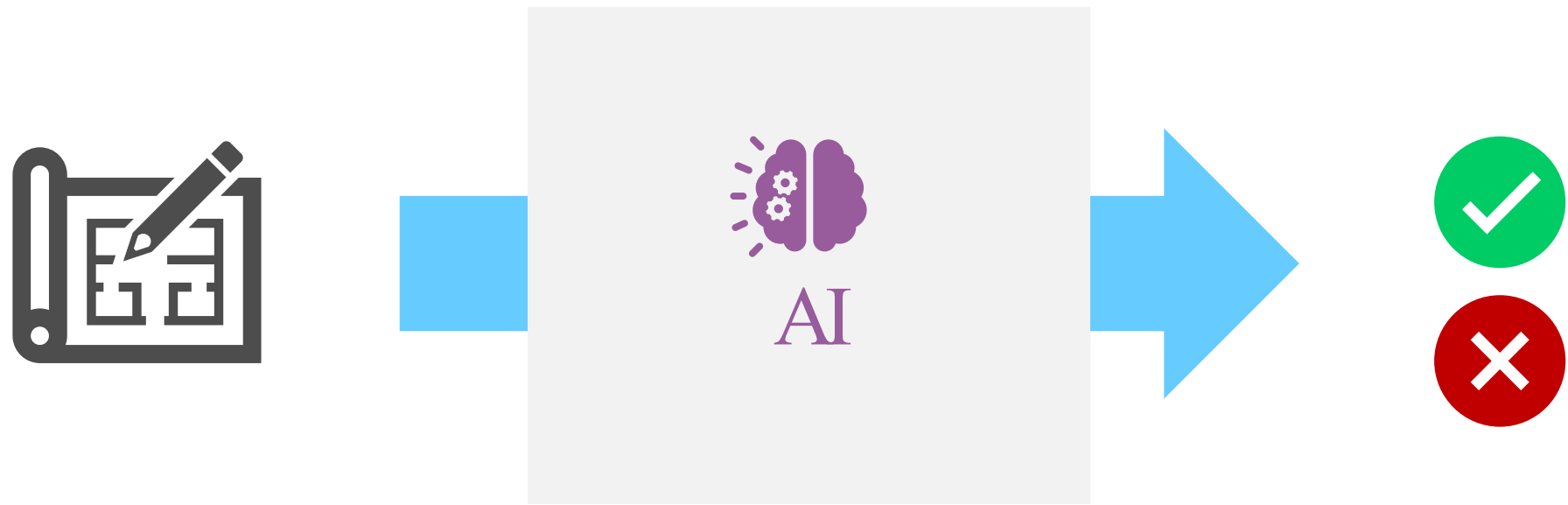


1500 drawing submissions per year





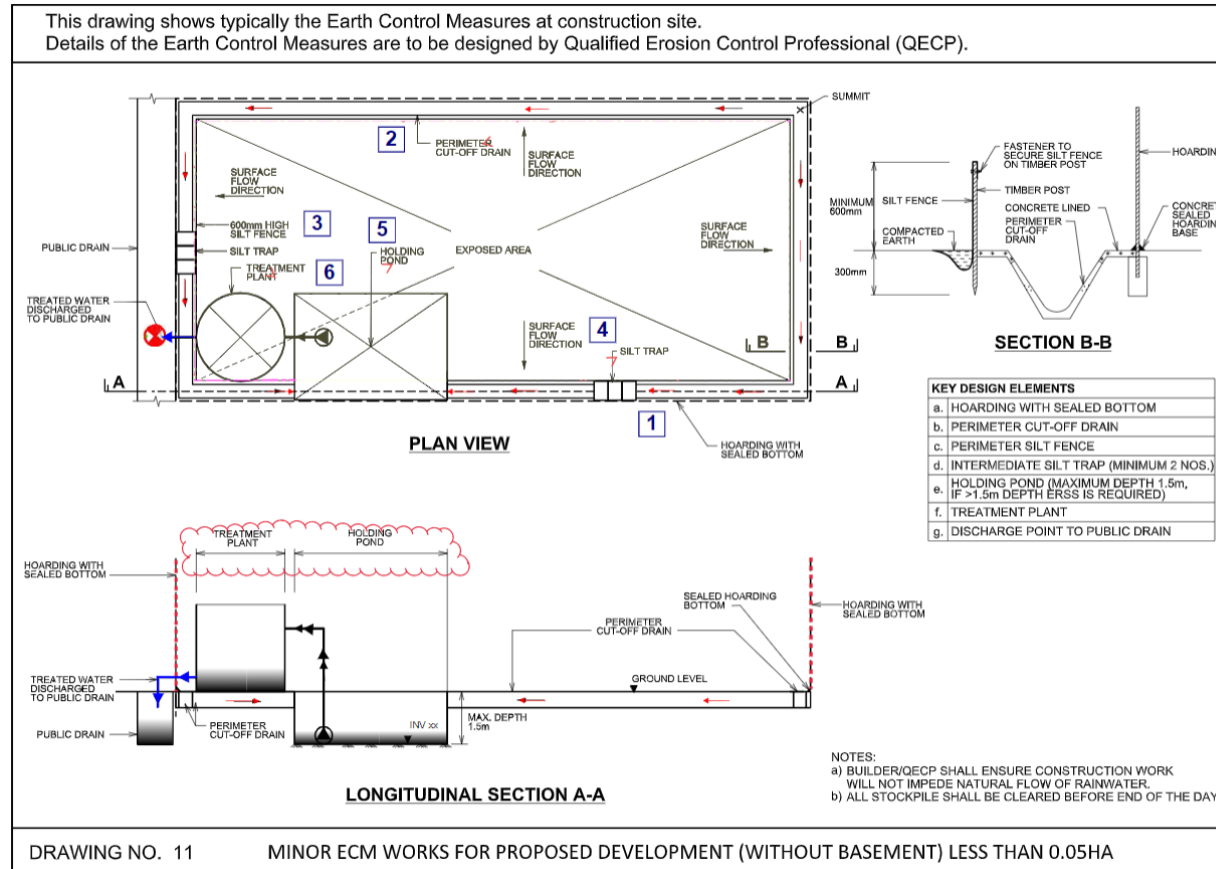
# Desired solution: Full automation with Machine Learning





# Expectations vs. Reality

# Expectation: Well-readable, standardized plans



# Reality: Huge variance & complexity in the data



Plan Complexity



Notation



Symbols



Naming



Information Conv.



Overlap

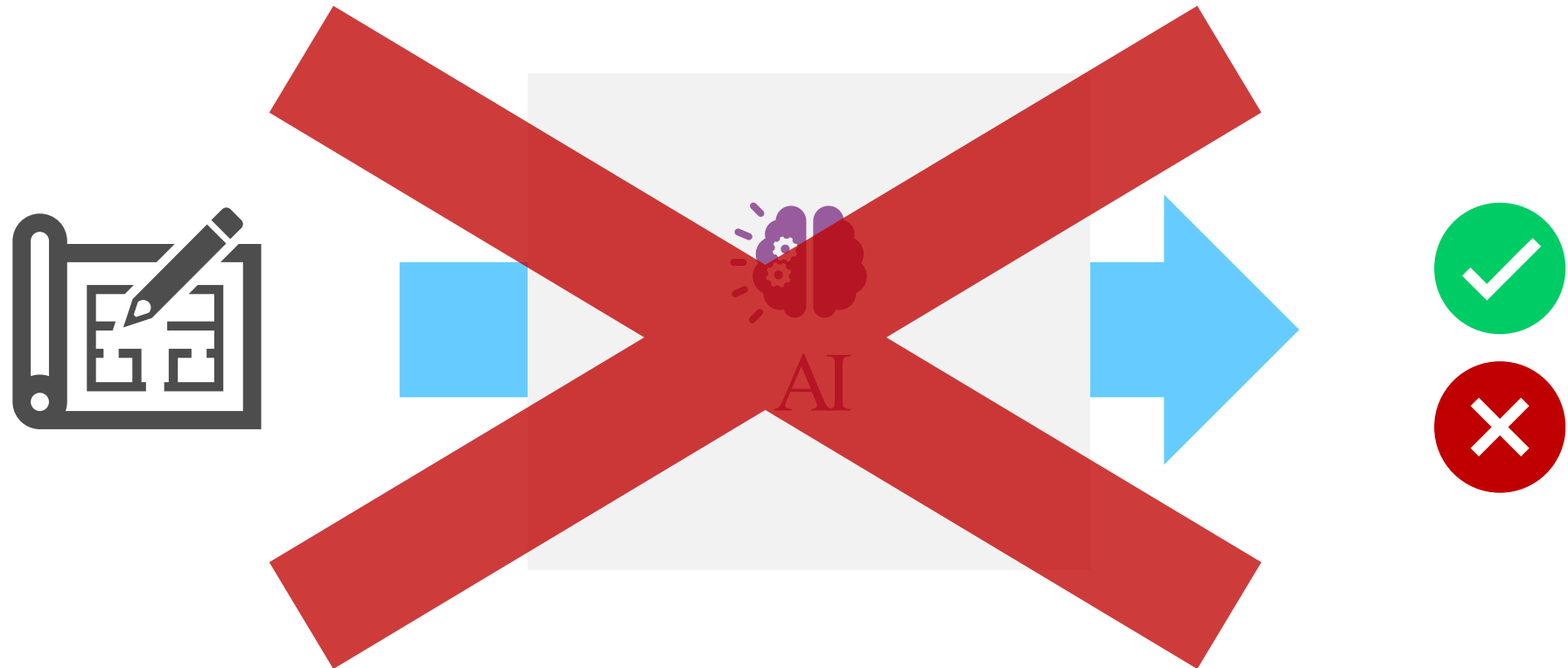


Low Resolution

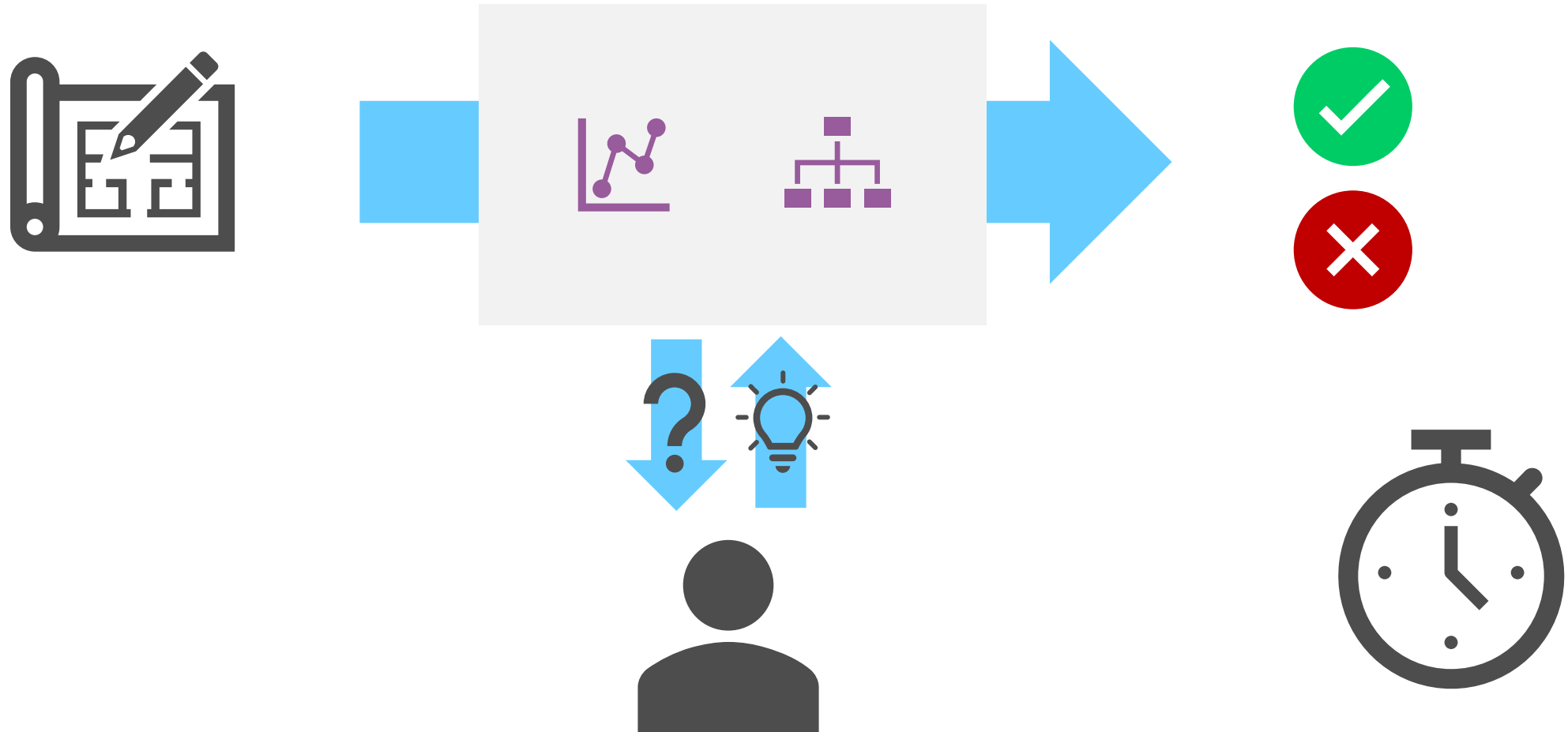


Spelling Errors

# Conclusion: Full automation not possible

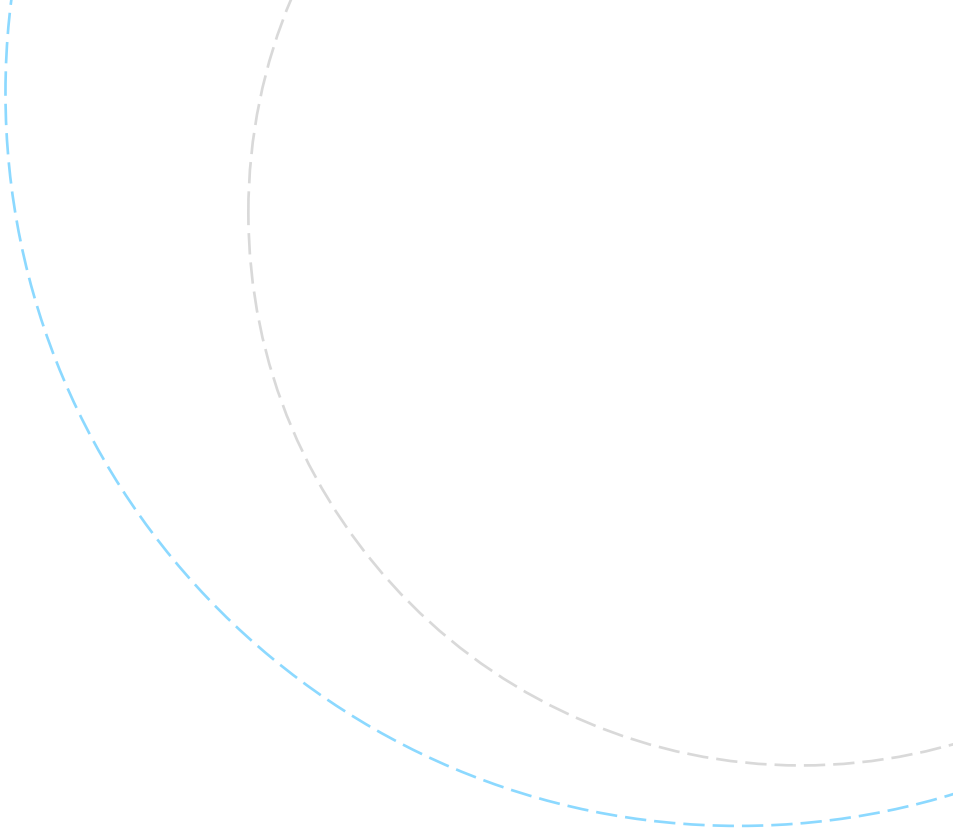


# Approach: Human in the loop

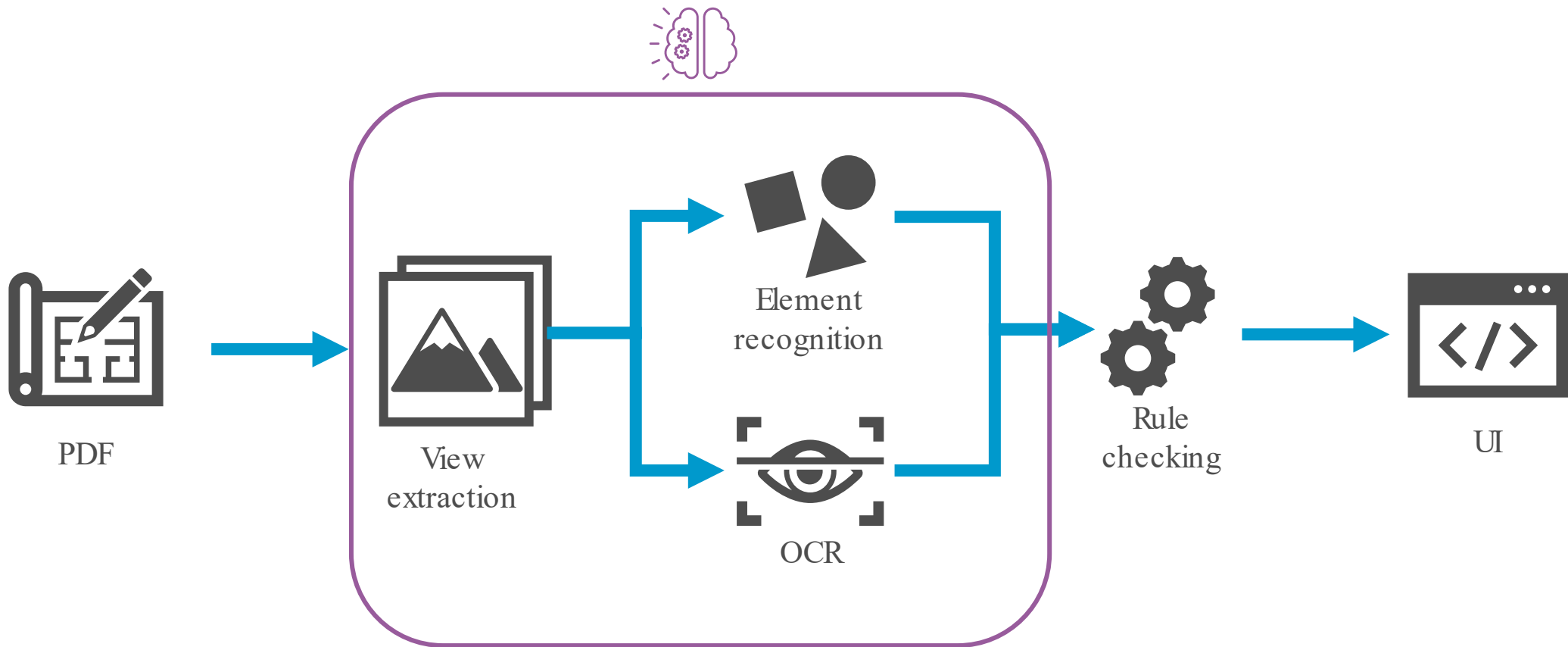




# Approach

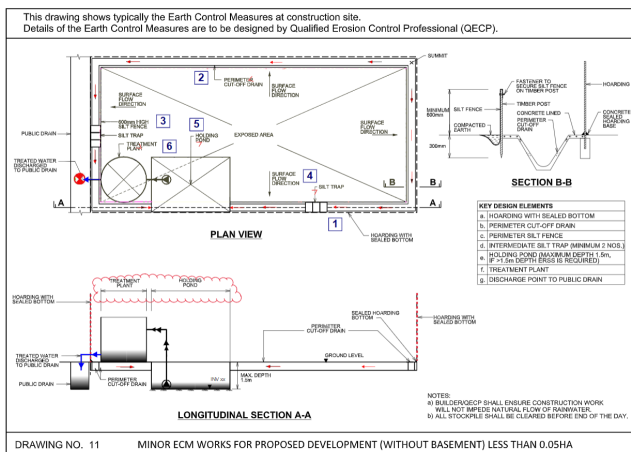


# Overview diagram

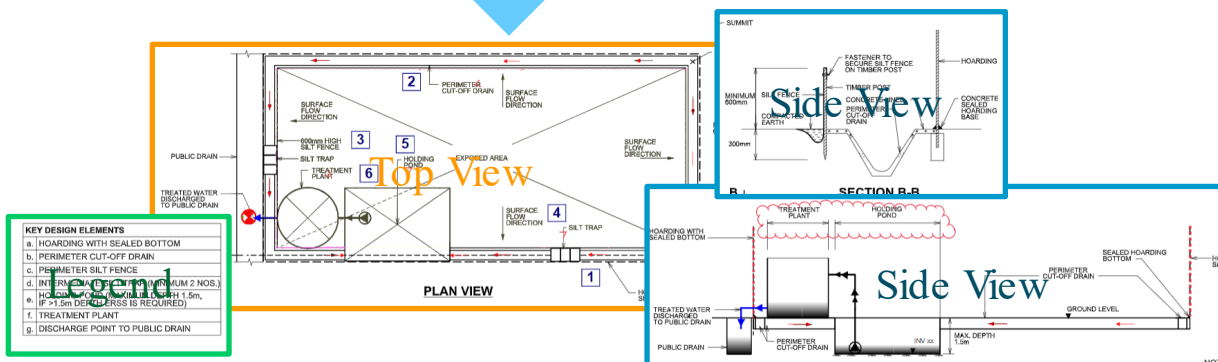




# View Extraction



View detection



Step 1:  
Annotate Data

We used VGG Image Annotator to label views. A total of 402 images (237 training + 165 testing) were prepared

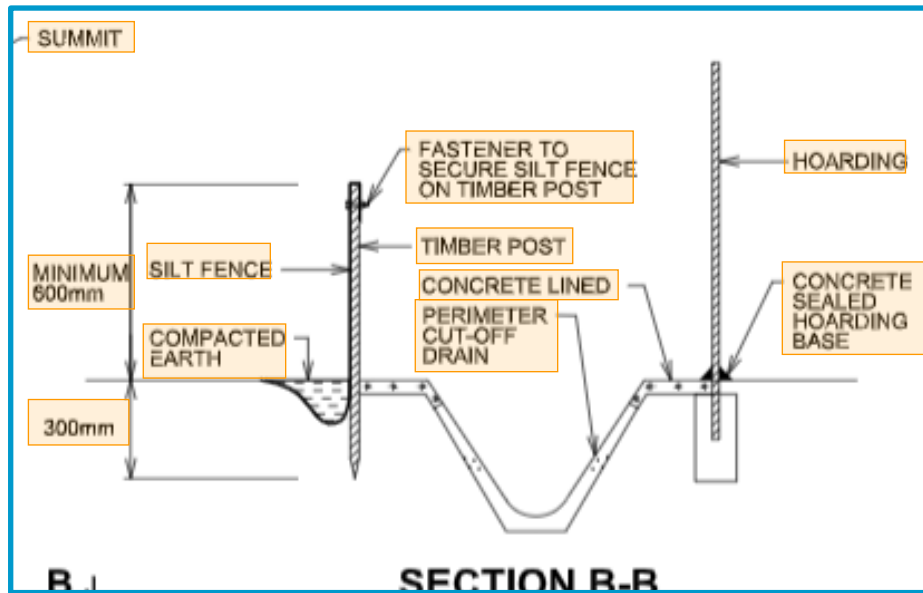
---

Step 2:  
Used an object detection model

**Detectron2**

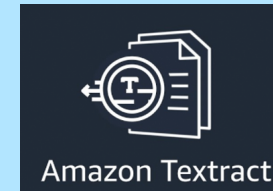
We used a mask-RCNN model in Detectron2

# OCR – Optical Character Recognition



## Step 1:

Used an OCR model to extract text from images



We used AWS Textract

## Step 2:

Postprocessing



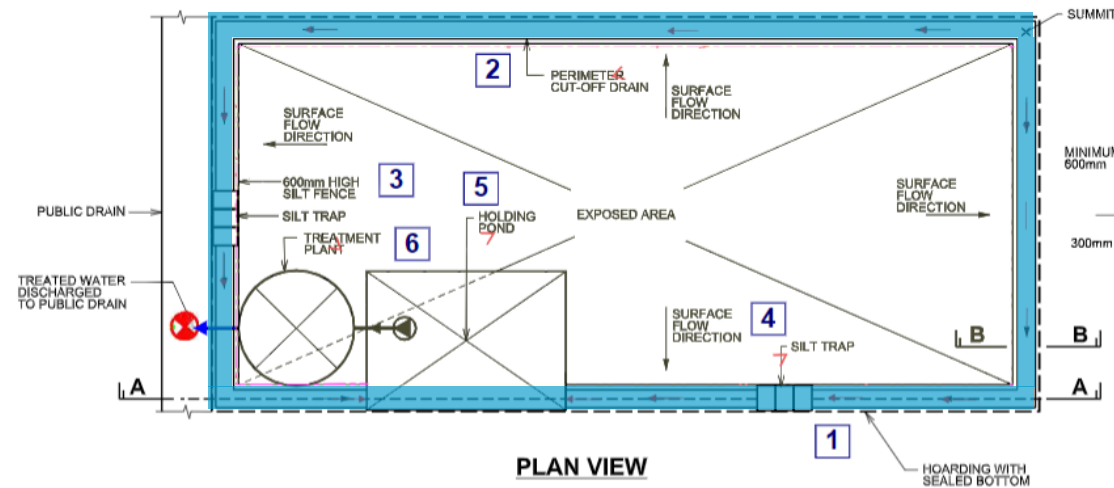
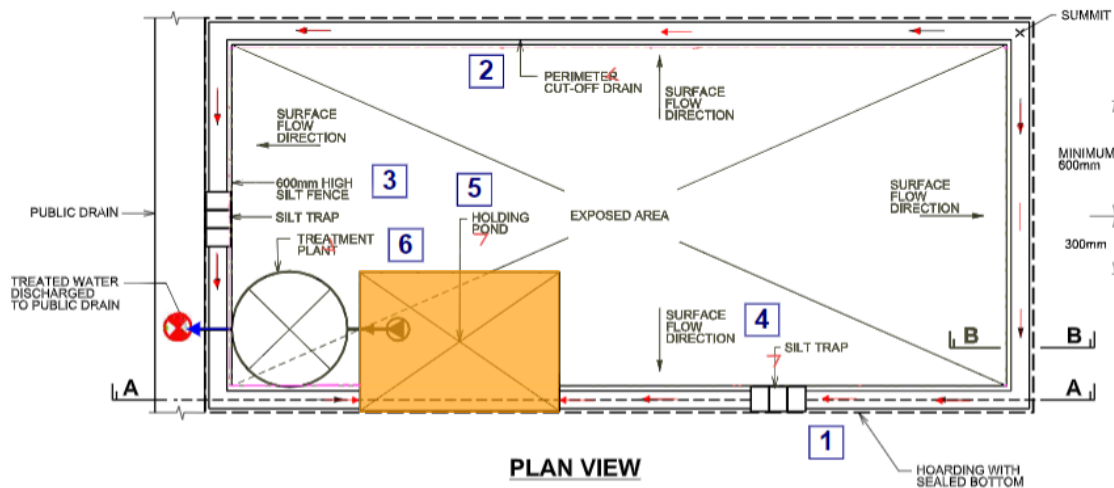
- We applied 3 postprocessing steps:
  - **#1, Paragraph grouping.** Textract recognizes single lines. Grouping based on distance and alignment
  - **#2, Spelling corrections.** E.g. “Hoarding” detected as “hoardino”. We check all words at hamming distance 1 that match a keyword.
  - **#3 Vertical Text recognition.** We rotate the whole image.

# Element Recognition

“things” vs “stuff”

**Things** Localized, countable elements

**Stuff:** Uncountable elements, spanning over an extended area



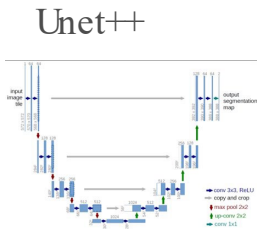
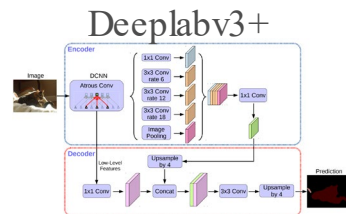
“Things” examples Cross section, discharge ponds, flow direction, silt traps, storage ponds, treatment plants

“Stuff” examples Cut-off drain, exposed area, hoarding, silt fence

Instance segmentation



Semantic segmentation



# Rule Types

Hoarding encompasses site

Position

Number of storage ponds equals number of treatment plants

Number

Storage pond volume is  $\geq 50\%$  run-off volume

Value

Arrow direction

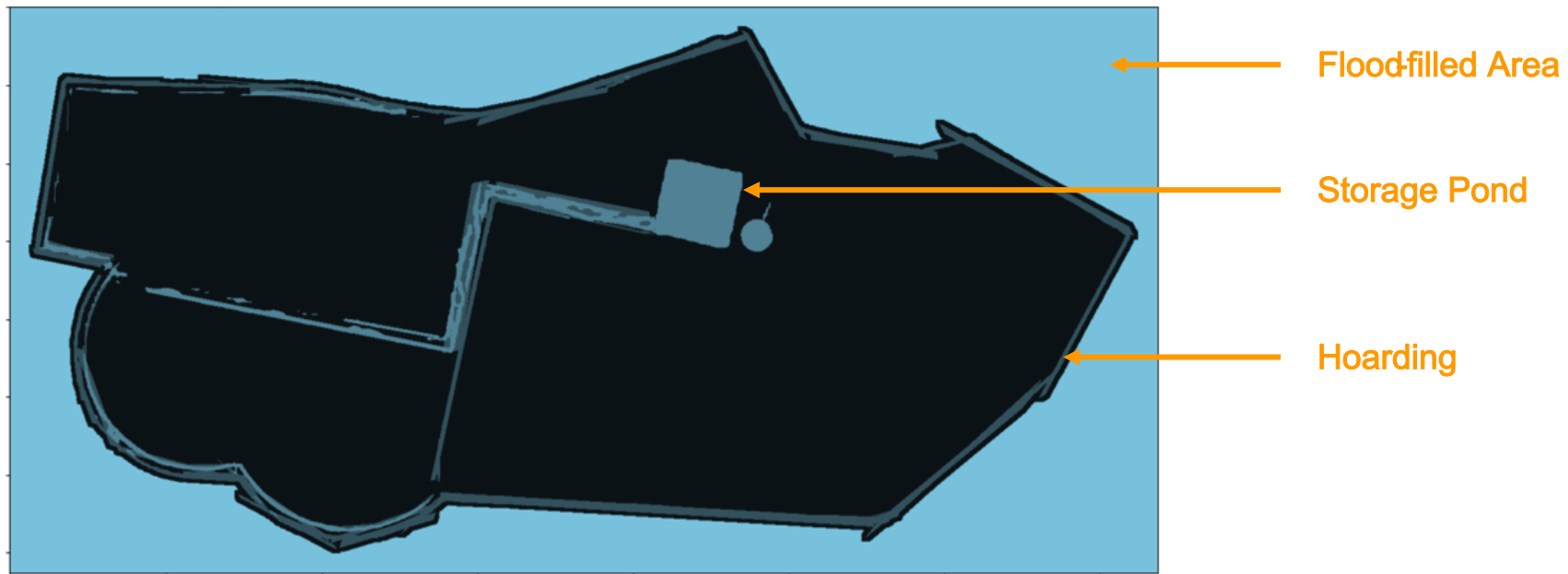
Connection

Height

Shape

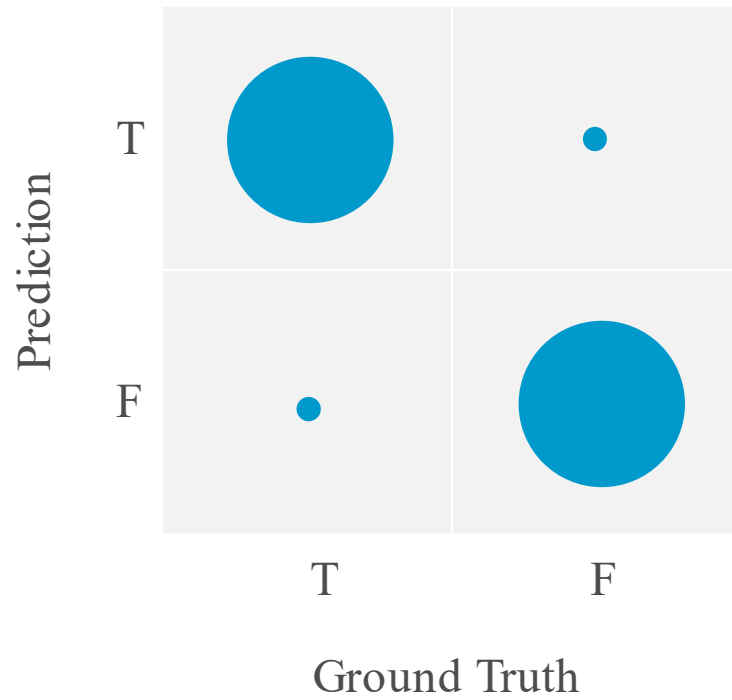
# Example Rule Checking

- Example of a rule: **Storage pond is within the hoarding boundary**
- Rule is of type position. To check it, the recognized hoarding and storage pond were used
- We flood-filled the outer area and made sure the storage pond was not touched

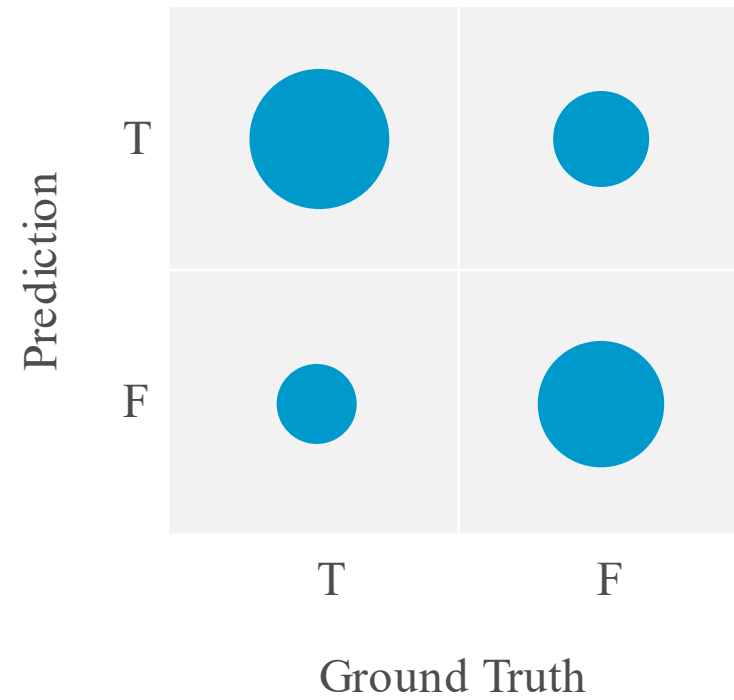


# Evaluation of whole System

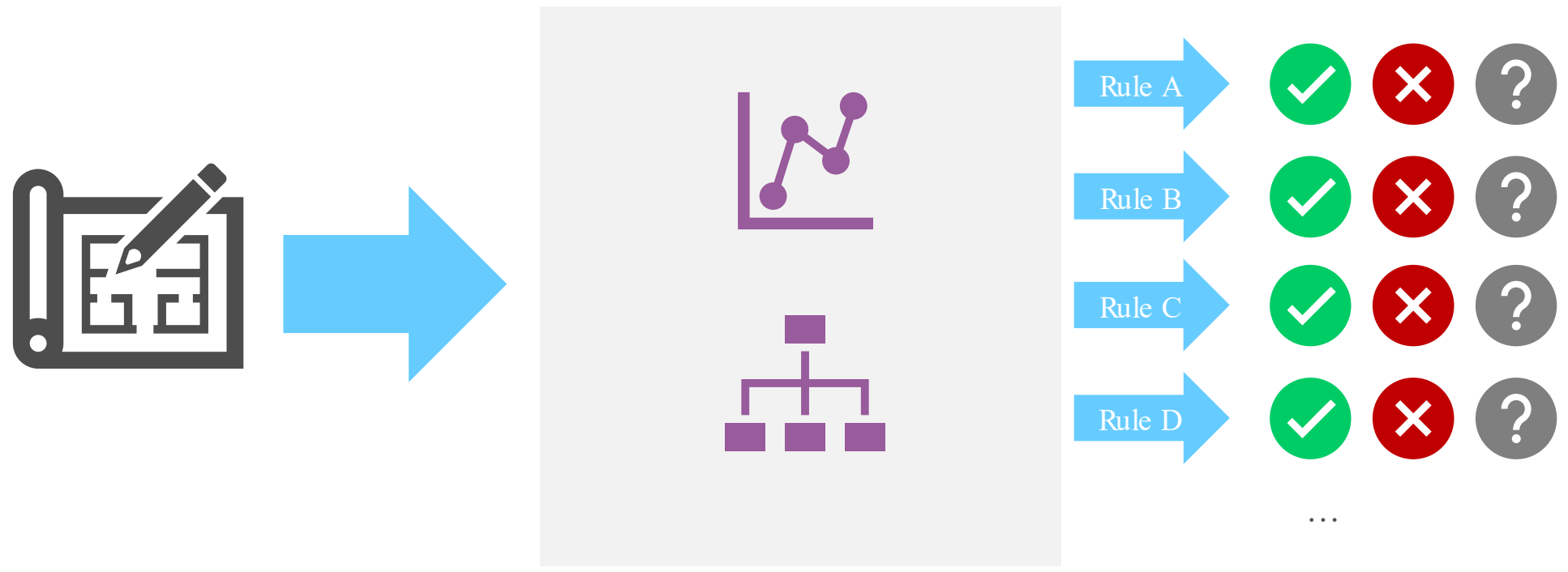
## Expectation












## Reality



# Human-in-the-loop to the rescue

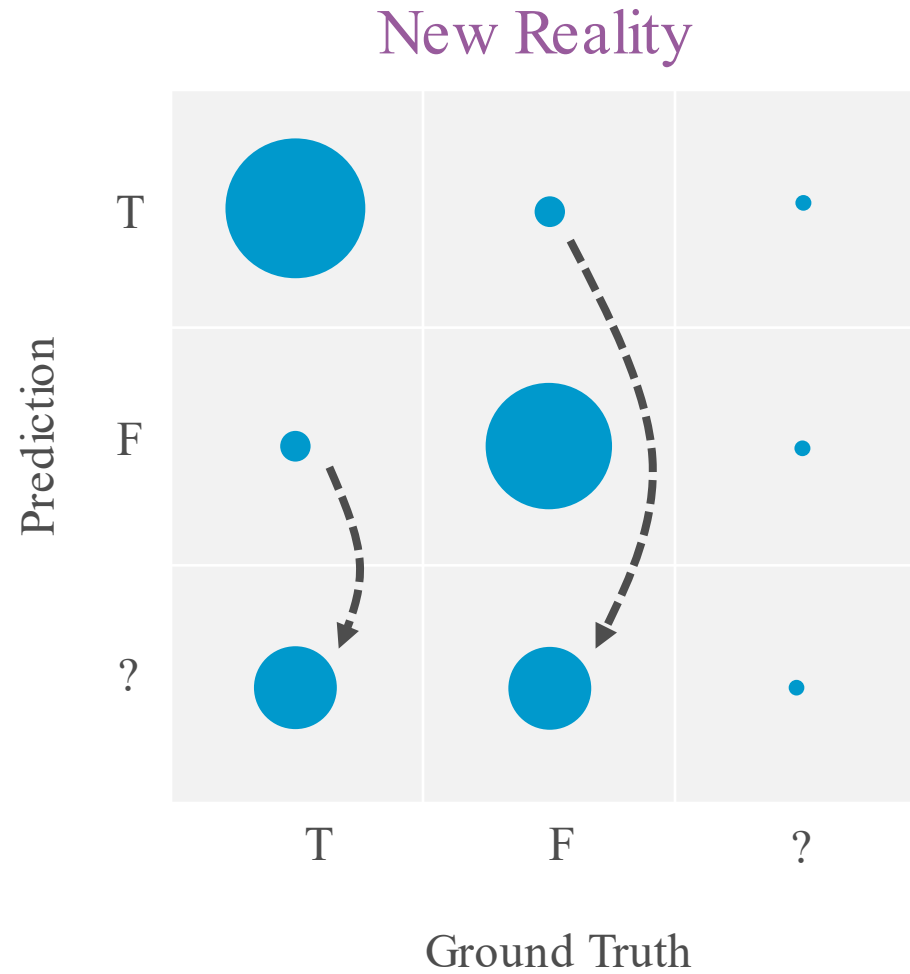


# No answer is better than a wrong answer

Prediction	T			
	F			
	?			
		T	F	?
		Ground Truth		



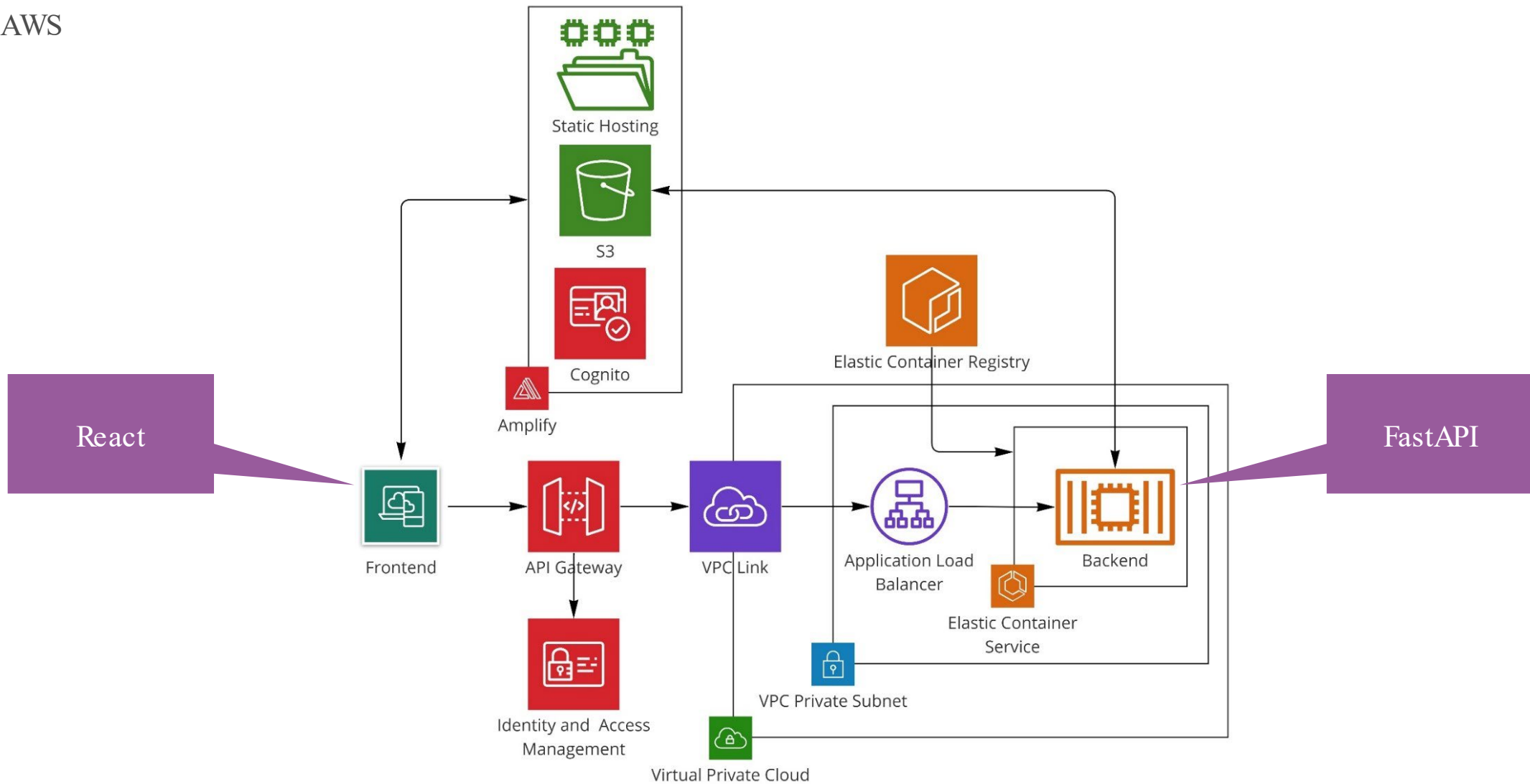
# Tune rules to be cautious





# UI & Deployment

# Architecture in AWS



PUB ECM Evaluator Prototype

ECM plan4 - Sanitised\_scan.pdf X

Rules

- ✓ Hoarding with sealed bottom.
- ✓ Hoarding encompassing site.
- ✓ Hoarding forms a closed loop.
- ✓ Hoarding along the site boundary.
- ✓ Exposed area within the hoarding line.
- ✓ Exposed area less than or equal to total site area.
- ✓ Unexposed area within the hoarding line.
- ✓ Unexposed area less than

status  
**ACCEPTED**

The screenshot displays the 'PUB ECM Evaluator Prototype' interface. At the top, a blue header contains the title 'PUB ECM Evaluator Prototype' on the left and 'ECM plan4 - Sanitised\_scan.pdf X' on the right. The main area is divided into three sections:

- Map:** A technical drawing of a site plan with various colored lines and shapes representing different elements. A red crosshair is visible on the map.
- Legend:** A vertical panel on the right side of the map, outlined in red, titled 'Legend' in red text. It lists various symbols and their corresponding descriptions, such as 'HOARDING WITH SEALED BOTTOM', 'HOARDING ENCOMPASSING SITE', and 'HOARDING ALONG THE SITE BOUNDARY'.
- Rules Panel:** A vertical panel on the far right titled 'Rules'. It contains a list of ten rules, each with a green checkmark icon and a dropdown arrow. The rules are:
  - Hoarding with sealed bottom.
  - Hoarding encompassing site.
  - Hoarding forms a closed loop.
  - Hoarding along the site boundary.
  - Exposed area within the hoarding line.
  - Exposed area less than or equal to total site area.
  - Unexposed area within the hoarding line.
  - Unexposed area less than

At the bottom of the Rules panel, there is a 'status' field with the text 'ACCEPTED' in white on a dark green background. On the left side of the map, there are three blue buttons: 'Views' (with a diamond icon), 'Elements' (with a grid icon), and 'Text' (with a text icon). Below the 'Views' button, there are four smaller buttons: 'All', 'Legend', 'Top', and 'Side'.

PUB ECM Evaluator Prototype

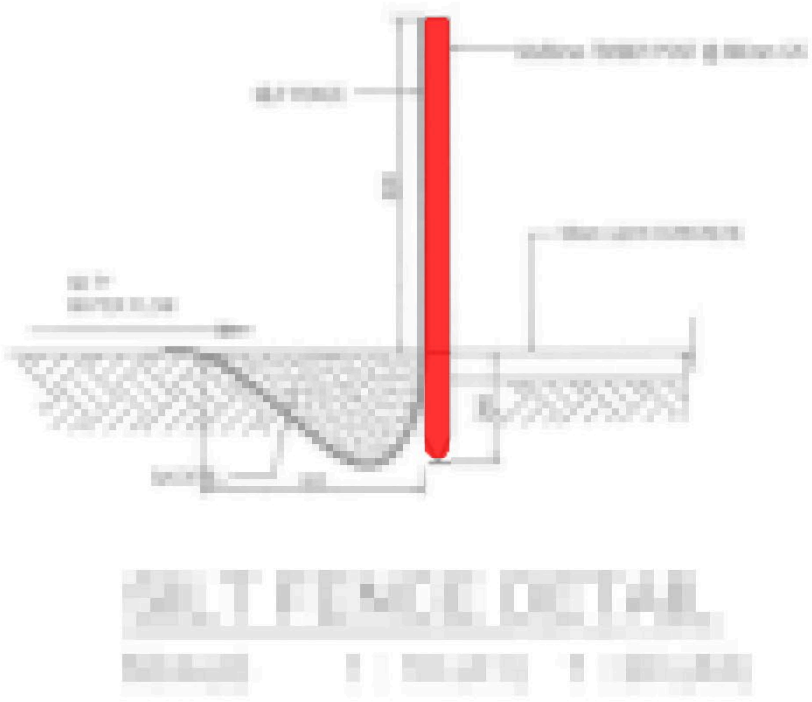
ECM plan4 - Sanitised\_scan.pdf X

- cut\_off\_drain
- treatment\_plant
- storage\_pond
- discharge\_point
- cross\_section
- silt\_trap
- concrete\_hump
- flow\_direction
- public\_drain
- ✓ timber\_post
- earth\_bund
- exposed\_area

Views

Elements

Text



Rules

- ✓ Hoarding with sealed bottom.
- ✓ Hoarding encompassing site.
- ✓ Hoarding forms a closed loop.
- ✓ Hoarding along the site boundary.
- ✓ Exposed area within the hoarding line.
- ✓ Exposed area less than or equal to total site area.
- ✓ Unexposed area within the hoarding line.
- ✓ Unexposed area less than

status

**ACCEPTED**

PUB ECM Evaluator Prototype

ECM plan4 - Sanitised\_scan.pdf

Views

Elements

Text

hoarding

exposed area

unexposed area

concrete hump

timber post

silt fence

flow direction

cut-off drain

silt trap

storage pond

treatment plant

discharge point

site boundary

search text  
run-off

< 1 / 2 >

Rules

- Hoarding with sealed bottom.
- Hoarding encompassing site.
- Hoarding forms a closed loop.
- Hoarding along the site boundary.
- Exposed area within the hoarding line.
- Exposed area less than or equal to total site area.
- Unexposed area within the hoarding line.
- Unexposed area less than

status

**ACCEPTED**

AREA EXPOSED AREA (m <sup>2</sup> )	4000	SEDIMENT YIELD (g/m <sup>2</sup> )
<b>RUNOFF COEFFICIENT</b>	0.45	TOTAL TREATMENT PLANT
<b>TOTAL RUNOFF VOLUME (m<sup>3</sup>)</b>	200	TOTAL HOLDING TANK CAP
		TOTAL CUT-OFF DRAIN CAP
		TOTAL STORAGE PROVIDED

PUB ECM Evaluator Prototype

ECM plan4 - Sanitised\_scan.pdf

**Rules**

- Hoarding with sealed bottom.
- Hoarding encompassing site.
- Hoarding forms a closed loop.
- Hoarding boundary...
- Exposed hoarding line.
- Exposed area less than or equal to total site area.
- Unexposed area within the hoarding line.
- Unexposed area less than...

status  
**REJECTED**





# Next Steps

# Next Steps

A full automation is not feasible if the data is not standardized.



Re-evaluate business case



Standardize submissions



Build MVP

Thanks!

Keeping Waterways Clean with ML

Paola Bianchi, Silvan Melchior | SDS 2024