a1829445, **Kelvin Ting** • a1808891, **Pham Gia "Alan" Nghi** • a1805827, **Kelin "Colin" Zhu**

Supervised by **Thomas Rowntree**

Project Introduction

Police officers receive textual description of criminals via the computers in their cars. Our system reads this textual description and converts it into an image to reduce the cognitive load of the officers.

Data Generation

Labels

Gender	Race
male	white
female	black
man	asian
woman	caucasian

Sentence Structures

Type1:
A/The race gender with hair_length hair_color hair, is wearing top, bottom and shoes, who also acc_verb accessory.

Type2:
A/The race gender in top, bottom and shoes, has hair_length hair_color hair. The person also acc_verb accessory.

Extra Sentences

Extra sentences F	Extra sentences B
The location is near the Royal Adelaide Hospital.	The person appears drug-affected.
There is a car crash.	The situation is very dangerous.

Input Training Data

Relative Feature Data

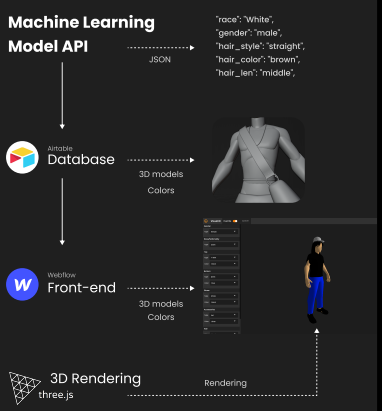
"race": "White",
"gender": "male",
"hair_style": "straight",
"hair_color": "brown",
"hair_len": "middle",
"top": "shirt",
"top_color": "black",
"bottom": "NA",
"bottom_color": "NA",
"footwear": "NA",
"footwear_color": "NA",
"accessory": "gun",
"accessory_color": "NA"

Synthetic Sentence Data

The location is near the Royal Adelaide Hospital. A White male, middle brown straight hair, a black shirt, is holding a gun. The situation is very dangerous.

Extra Sentence · Features · Sentence Structure

Image Generation



Machine Learning

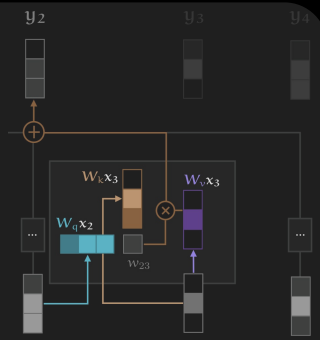
Transformer

Transformers are state of art machine learning architectures for NLP.

Self-attention mechanism that captures relationship between sequential elements that are far from each other.

These roles are often called the **query**, the **key** and the **value**.

RoBERTa consists of a simple stack of transformer blocks, of the type we described above but fined-tuned and optimized for NLP tasks.

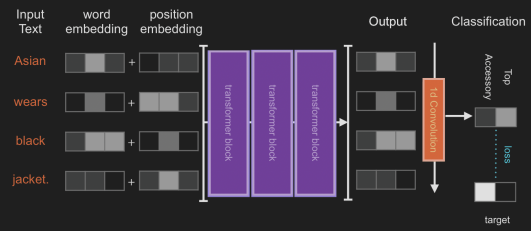


RoBERTa

The input text is tokenized and embedded according to its word and position.

Transformation blocks transform the embedding for capturing the sequence to sequence and parallel relationship from each other.

Finally, the embedding goes through a convolution layer for classification.



System Architecture

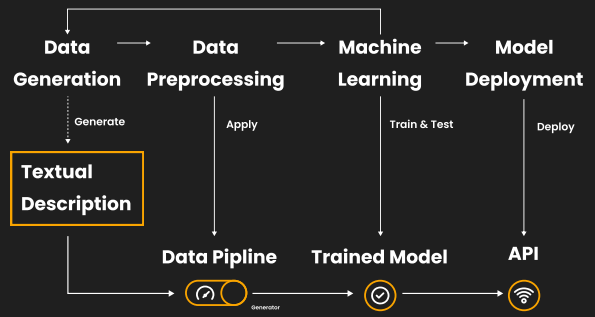


Image Generator

99% accuracy
1 second completion

Visual ID

Text to Perpetrator, visualising human description

Conclusion

Visual ID represents an innovation that can impact the police industry and other equivalent industry as a whole.

With more time in development, and more accurate data from real-life cases, the app can extend further to assist not just within the police force but also in other community service areas such as fashion design.

Experience **Visual ID**

Python Data Generator · Pytorch Machine Learning · Docker Model Deployment · Google Cloud Model Hosting · Webflow Front-end · Airtable Database · 3D Rendering three.js