Multimodal Machine Unlearning Thesis Proposal

giovanni.stilo@univaq.it, andrea.dangelo6@graduate.univaq.it

Machine Unlearning Basics

 Machine Unlearning is the task of removing undesired data from the training set of a Machine Learning model.

- Read more:
- <u>https://arxiv.org/pdf/</u>
 <u>2306.03558</u>
- <u>https://arxiv.org/pdf/2209.02299</u>



The process starts with a dataset (1), which is split into a Train Set and a Forget Set (2). The Train Set is used to train the Original Model (M), while the Forget Set contains samples to be unlearned. An Unlearning Model (3) is applied to remove the influence of the Forget Set, resulting in an Unlearned Model. A Retain Set is used to train a Gold Model (GM) for comparison. Finally, Evaluation Metrics (4) assess the effectiveness of unlearning by comparing the Unlearned Model to the Gold Model.

PlayMyData Starting Point

- PlayMyData is a curated dataset composed of 99.864 multi-platform video games.
- It includes screenshots, video URLs, tabular information, and textual descriptions.
- Read more:
- <u>link</u>



Thesis Main points

- Study the literature on Multimodal Machine Unlearning
- Extend PlayMyData to include more features
- Publish the new Dataset on TorchVision for new research opportunities
- Benchmark current Unlearning algorithms on multimodal datasets