Semester Project Minor Data Sciences

Project Report - June 6, 2025

TOWARDS A DATABASE OF CIVILIAN DEATHS IN THE CONTEXT OF POLICE INTERVENTIONS

Student: *Stefanie Helfenstein*
Supervisor: *Jérome Baudry*
Supervisor Border Forenscis: *Stanislas Michel*

During the semester in which this project was developed, at least two individuals died in the context of police interventions. This work is dedicated to their memory.

Contents

1 Introduction

1.1 Context

To date, there is no centralized database recording deaths occurring during police interventions in Switzerland. In contrast, several other countries maintain such statistics, whether through official channels or independent initiatives. For example, in France, data is officially collected by the IGPN (Inspection Générale de la Police Nationale) [6], as well as independently by the media outlet Basta [3]. In England, the organization Inquest [5] gathers this information, while in the United States, projects such as 'Mapping Police Violence' [9] and the Washington Post [10] provide similar documentation.

1.2 Problem

In Switzerland, various organizations, such as Augenauf [2] and Amnesty International [1], have attempted to collect such information locally and on a partial basis. However, there is currently no comprehensive database documenting deaths resulting from police interventions. This lack of systematic data collection hinders efforts to shed light on critical issues, particularly in a context where the UN Working Group of Experts on People of African Descent has raised concerns about structural anti-Black racism and impunity in Switzerland [11].

1.3 Border Forensics

This semester projects collaborates with Border Forensics for a project about deaths occuring in context of police interventions in Switzerland. Border Forensics is an NGO based in Geneva that is conducting investigations into practices of border violence perpetrated by states, police forces, militaries, and corporations [4].

Border Forensics provided a manually compiled database documenting deaths occurring in the context of police interventions in Switzerland. This dataset served as a reference point and ground truth to evaluate the performance of our code. It was updated with newly identified cases and used throughout several stages of the analysis, as detailed in the following sections.

2 Goal

As part of a semester project supervised by Prof. Baudry from the Laboratory for the History of Science and Technology (LHST) and the NGO Border Forensics, our goal is to develop an informative database of deaths occurring in the context of police interventions in Switzerland. Drawing on the Archives from the newspaper 'Le Temps' [8], the project aims to establish a methodology for an automated and as exhaustive as possible identification of cases of civilan deaths resulting from police interventions.

One of the main challenges encountered at the beginning of the project was defining a "death occurring in the context of a police intervention." What types of cases should be included? In France, the IGPN (Inspection Générale de la Police Nationale) maintains an official record of individuals who died in such contexts. This includes deaths resulting from firearm use, deaths related to the physical or mental health conditions (including addictions) of the deceased, deaths caused by the use of physical force during arrest, deaths influenced by the behavior of the individual, and fatalities resulting from traffic accidents—provided that police presence was involved in each case [6]. Nuances important to take into account are cases such as distinguishing between deaths in police custody (included) and deaths in prison (excluded, as they fall under the responsibility of the prison system rather than the police).

We therefore defined the cases to be included in our dataset on the definition that was as well used by the journal Basta! [3]:

any civilan person killed or deceased as a result of an interaction with a police or gendarmerie officer, whatever the circumstances that led to this intervention, whatever the profile of the persons killed, and whatever the nature and seriousness of the offence of which they may be suspecteds.

The purpose of this database is to better understand how police interact with civilians and the outcomes of these interactions. Consequently, we limited the scope of the database to civilian deaths only. The circumstances surrounding each death were described using metadata that we aimed to extract throughout the project.

2.1 Metadata

The final metadata we seek to extract falls within the following categories. Given the nature of the sources, expanding these categories would be relatively simple.

Date	*Number of deaths*
Name, age, gender	*Region of origin*
Nationality	*Cause of death*
Source	*Place of death*
Latitude	*Longitude*
Death type	*Language of source*

2.2 Reusable code

Given the collaboration with the NGO, the a secondary goal of the project was as well to create a code that is reusable for the NGO, not depending therefore of the huge computational power provided to us by clusters of the EPFL. However, for time reasons and experimental try-outs, the project still worked with the cluster until finding the final version of the code. This version then needed to work on the commercial computers with their computational power.

2.3 Wording

Another question that needed to be adressed was the one of wording. While the IGPN of France used catégories such as *'deaths during traffic accidents'*, Border Forensics used the wording of *'traque'*. The words used already project a judgement of the cases on the lecture (in the first one, the fault was of no one and it was just an accident, the second one projects more an image of violence). Lectors of the dataset will need to be aware of those difference and take it into account while plotting the graphs. Personnally, for the project I choose to go with the categories choosen by Border Forensics.

2.4 Respect of Privacy

The dataset collects data of individual people, therefore the question of privacy is of high importance. As for the visualizations it hasn't been adressed yet how to display personal data nor the family and friends have bseen consulted, the final dataset rests confidential and only large tendencies will be displayed.

3 Methodology

To carry out the project, two main phases were adressed. The first involved discussions and interviews with experts in the field, aimed at deepening my understanding of the topic and integrating the insights gained into the project's approach. In the second phase, the code was developed based on the acquired knowledge.

3.1 Interviews

During the exploration phase, diverse interviews have been conducted with experts in the topic, such as a lawyer working on a case regarding death in context of a police intervention, a journalist recording with Basta! such a database for France as well as an Human rights expert of Amnesty International working on police violence in Switzerland. These discussions allowed to highlight interessting points regarding the work on deaths occuring during police interventions.

Journalist of Basta

Basta! has been collecting cases of deaths in the context of police interventions long before the IGPN (Inspection générale de la police) started collecting the cases. Since 2018 the IGPN (Inspection générale de la police) is collecting their own data, since 2020 they are publishing the details about what cases they took into account. The IGGN (Inspection générale de la gendarmerie nationale) was doing the same in 2021. This allows a comparison between the collected data of Basta! and the state collection.

Another point raised by the journalist was the importance of involving the families and friends of the deceased in the decision-making process regarding how the data should be used and represented. Their participation would be essential in guiding the next steps, particularly in the ethical use and contextualization of the dataset.

Expert of Amnesty

Amnesty International is not engaged in the continuous collection of this type of data but has conducted several targeted studies on police violence. The discussion emphasized the importance of drawing from diverse sources, including newspapers, social media, and legal channels such as lawyers. Much of their research relied on leveraging their existing network. They also offered to connect us with several of their contacts, including a lawyer who has represented families of individuals who died during police interventions.

Lawyers

The human rights expert from Amnesty connected us with lawyers working on cases that fall under the category of deaths occurring in the context of police interventions. Although they couldn't help in identifying new cases to add to the database, they also cautioned against relying on legal documents for expanding the dataset, a possibility that crossed our minds when working on the database. There is no single legal category encompassing all such cases, many case files are not publicly accessible, and several deaths never reach the courtroom at all. Instead, they provided insights into the legal follow-up of such cases, including court procedures and developments. While this information was less directly relevant to the current project, the discussion opened up valuable perspectives for future work, such as analyzing the legal trajectories of the cases already included in the database.

3.2 Code

As a first step in the coding approach, a database of articles was compiled. Using two distinct pipelines, it was then attempted to extract the most accurate possible list of articles reporting on cases of deaths occurring in the context of police interventions. These pipelines will be presented in the following section.

3.2.1 Sources

As a source for the coding approach, I began with the newspaper Le Temps, which offers well-structured online archives and covers the time period (1998–2025). However, it is worth noting that the Ground Truth document from Border Forensics spans a broader period (1993–2025), meaning that five years are omitted by relying only on 'Le Temps' as a source. The newspaper's archives are organized into three hierarchical levels: first by year and month, then by article titles, and finally by full article content. This structure was used to begin the analysis with a reduced amount of textual data, gradually adding more detailed information at each stage of the process.

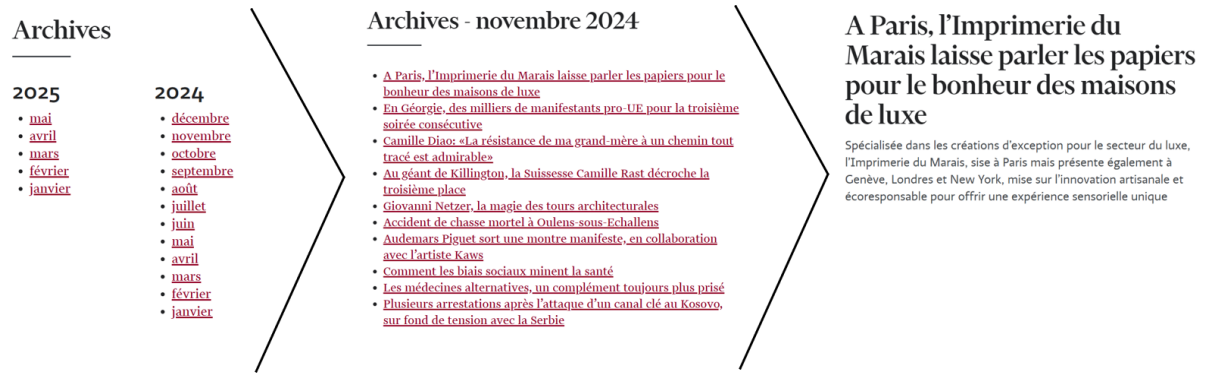

Figure 1: Different levels of online archives of 'Le Temps'

3.2.2 Pipeline

As the database contains a huge amount of articles and NGO's are not working with computers of high computational power, a pipeline was chosen to reduce the number of data working with through two steps. Another reason for choosing this pipeline was the fact, that scraping such a huge amount of data was sometimes overstraining the system:

```
1    requests.get()
```

request) was not able to handle that much requests at once. By reducing the size at each step before adding new data, we were able to minimize the number of requests made at once, which allowed the system to function properly. We implemented two distinct pipelines: the first tested a more generalizable approach that should be bestter adaptable to other newspapers, while the second is adapted specifically to 'Le Temps', incorporating the newspaper's internal article categorization structure.

The main difference between the two pipelines lies in how they handle location data. This second pipeline reduces the dataset by extracting mentioned locations and mapping them onto a world atlas. As a result, the integration of the Lead Posts occurred at a different stage in the process. However, the keyword reduction and model analysis steps remained identical in both pipelines and are detailed below.

Pipeline 1

The following outlines the general structure of the first pipeline. In this setup, the Lead Post data is integrated after the Keyword Reduction Step, but before the dataset is filtered based on location.

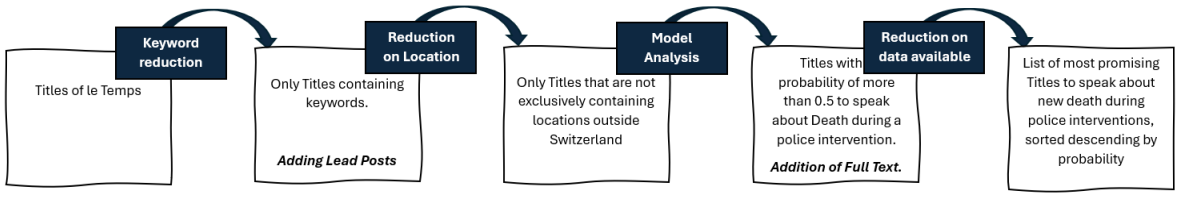

Figure 2: Pipeline 1 for reduction of the size of the dataset and developement of final list

Pipeline 2

The second pipeline applied the same keyword filtering step. However, it uses the structure of the website of the Archives of 'Le Temps', where articles related to events in Switzerland are categorized under URLs beginning with "https://www.letemps.ch/suisse/". Thanks to this straightforward filtering method, the location-based reduction was performed before the keyword filtering step. The Lead Post data was then added later in the process, just prior to the model analysis. This resulted therefore in the following pipeline:

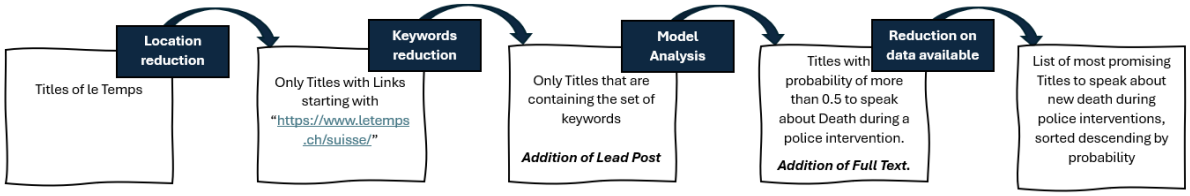

Figure 3: Pipeline 2 for reduction of the size of the dataset and developement of final list, applicable specifically on Titles of 'Le Temps'

The methods applied at each stage of the pipeline are detailed below.

3.2.3 Filter based on keywords

This keyword list used in the Keyword Reduction Step was constructed in three parts. Using the Border Forensics Ground Truth database as a reference, the titles and lead posts were extracted from all french sources of the gathered cases. From this textual data, words were selected that had a similarity score above 0.4 with the word 'police' or above 0.5 with the word 'death'. Due to confidentiality, the Border Forensics database is not included in the GitHub repository

```
nlp = spacy.load("fr_core_news_md")

police_ref = nlp("police")
nlp(word).similarity(police_ref) > 0.4

mort_ref = nlp("mort")
nlp(word).similarity(mort_ref) > 0.5

```

```
lst_handmade = ['tuer', 'cellule', 'arrest', 'officier',  'garde à vue',
     'balles', 'tir', 'coup de feu', 'acquitt', 'légitime défense', '
    classement', 'plaquage ventral', 'contr le  d identit é', 'détention
    ', 'forces de  l ordre ']
```

To this two lists, a handmade list of additional words was added that is based on the discussion with the different experts working in this domain. A list of 48 words resulted. With these three criteria, a list was created based on which the titles were filtered. Only the titles were kept, that contained at least one of the keywords. This list can be updated continuously, if we see that sone of the words are lacking.

```
Set = {'administratif', 'interpellation', 'procureur', 'classement', '
    tir', 'agent', 'défense', 'mourir', 'décès', 'accusation', 'mort', '
    tuer', 'garde à vue', ...}
```

To the extracted list of titles, the lead posts of each of the articles were added in order to have more textual data to analyze. However, the proportion of articles not having lead posts was low (Pipeline 1: 408/35553, Pipeline 2: 47/10394), therefore it was decided to ignore it and take for this articles only the titles.

3.2.4 Extraction based on Location

The next reduction of the datase based on the locations that are mentioned in the text is only used for Pipeline 1.

Pipeline 2 uses the caracteristics of the Links of the newspaper 'Le Temps', that are already classiying the titles in either 'monde' (world) or 'suisse (Switzerland).

```
df_CH = df[df["Link"].str.startswith("https://www.letemps.ch/suisse/")]
```

For the more general approache of Pipeline 1, three classifications have been decided:

- 0 if only locations outside of Switzerland have been found.
- 1 if a location situated in Switzerland has been found in the text
- 2 if no location could have been found in the text

First, the locations were extracted from the Titles and Lead Posts texts with help of the code:

```
def extraire_lieu(texte):
    doc = nlp(texte)
    lieux = [ent.text for ent in doc.ents if ent.label_ == "LOC"]
    return lieux


geolocator = Nominatim(user_agent="geo_checker")
location = geolocator.geocode(lieu, country_codes= "CH", addressdetails=
    True, timeout=7)
```

Two sets of extracted locations were created—one for places within Switzerland and one for locations outside of it. For Switzerland, it was checked if the extracted locations belonged to a geographic entities inside of Switzerland, that belongs to one of the following categories:

```
valid_categories = { "village", "town", "municipality", "city", "county"
    , "district", "state", "region", "state", "country" }
```

The classification of locations outside Switzerland was included because the location extraction algorithm sometimes identified terms that are not actual geographic entities as locations, such

as 'Bible' or 'Croque-mort'. Our goal was to exclude only those titles that could be reliably attributed to another country, not those containing falsely classified terms like 'Croque-mort.' To ensure accuracy, each extracted term was re-evaluated to confirm it represented a genuine geographic entity. However, the smallest category, *'village'* was excluded from this verification step, since nearly every place name can correspond to some village somewhere in the world. Yet, it is unlikely that such village names would be used in Swiss newspapers.

After having identified the georgaphical entities lying inside and outside of switzerland, a classification formula based on these sets was used, and each title was categorized into one of the three previously defined categories.

```
def classify_lieux(lst):
    if not lst:
        return 2

    def get_words(text):
        clean = normalize_str(text)
        parts = re.split(r"[   ']", clean)
        return [word for part in parts for word in part.split()]

    words = [normalize_str(word) for item in lst for word in get_words(
        item)]

    if any(word in suisse_list for word in words):
        return 1
    elif all(word in not_suisse_list for word in words):
        return 0
    else:
        return 2
```

At the end, every title attributed to the category 0 are excluded of the pipeline. Titles where no locations have been found or at least one location that can be attributed to Switzerland are kept.

This step allowed to reduce the dataset to a size of 34324 articles, meaning that the reduction through this step wasn't that important.

text	lieux	Classification
Chronologie de quatre ans et demi d'enquête – 2 septembre 1	['Halifax']	0
Frontières (Dominic Nahr / Magnum)	['Frontières']	2
Immigration «de masse»: tirer les leçons de Schwarzenbach (C	['Immigration', 'Université de Neu	1
De l'émergence de la police au XVIIIe siècle à son retrait des qu	['XVIIIe', 'Genève']	1
«La violence en Algérie a une raison. Elle vient du coup d'Etat (	['Algérie', 'Genève', 'Suisse']	1
La citation. Le vote militaire n'est pas gagné pour Bush "Les so	[]	2
Escalade des prix en Turquie: «Si ça continue, je ne sais pas ce	['Turquie']	2
Une rentrée visionnaire «1984», «La Servante écarlate», les dy	[]	2
L'analyse militaire «A défaut de lancer contre les forces serbes	[]	2
Conte de la violence économique au pays du brut «A Most Vio	[]	2
Le parquet de Lille classe l'enquête pour viol contre DSK «A pa	['de Lille']	2
Ne pas mourir au matin «A partir du moment où, après avoir v	[]	2
Le «sacré bouquin» d'Eluard et de Miró exposé à Montricher «	['Montricher']	1

Figure 4: List of articles with the identified places and the corresponding classification.

3.2.5 Language Model

The goal was to develop a language model capable of classifying article titles into the category "deaths in the context of police interventions" or not. Initially, the idea was also to experiment with separate classifications for the categories "police" and "death", to assess whether breaking them apart would improve performance. However, in the end, only the category "deaths in the context of police interventions" was used for evaluation. Despite this, the model was trained to recognize all three categories.

Test-Train split - Finetuning of model
For the finetuning of the language model, a manually tagged list of titles needed to be created. This list was combined through two different sources:

- A List of titles after the Step *Keywords reduction* has been taken.
- A List of the cases collected by Border Forensics and the corresponding Title in Newspaper, in order to create the Tags about *Death in the context of police intervention*

The list of titles was then manually tagged. At the end, a training dataset of in total 1122 Titles was obtained, in which the distribution was as following:

No tag	805
Death during police intervention	96
Death	217
Police	201
Death but no police	121
Police but no death	105
Total	**1122**

Table 1: Distribution of tags per categorie in the test-train split

For Splitting the dataset, the *MultilabelStratifiedShuffleSplit* was chosen, as it splits the dataset in a test and train set taking into account the distribution of the different lables of an unbalanced dataset, as it is the case in our dataset.

```
msss = MultilabelStratifiedShuffleSplit(n_splits=1, test_size=0.2,
    random_state=42)
```

As all of the Titles were taken in French, the model to train was choosen to be a CamemBERT model, that is specialized on the French language. It was opted for training the model on a multiclassification on the categories *'Death', 'Police', 'Death in the context of a Police intervention'*. However, as the category that is really interesting to us is mainly *'Death in the context of a Police intervention'*, this category was attributed a 3x the weight of the other categories.

Training
Then the training arguments were defined and a trainer was trained. We evaluated the training with the metric evaluation loss, which will stop training if the evaluation loss doesn't improve twice in a row.

```
early_stopping_callback = EarlyStoppingCallback(early_stopping_patience
    =2)
```

With all this parameters, a camemBERT model has been finetuned on the three categories *'Police', 'Death', 'Death in the context of a police intervention'*.

Evaluation & Use of model
With help of the test set, the model has been evaluated on its performance. As for the confusion matrix, this is giving us the following:

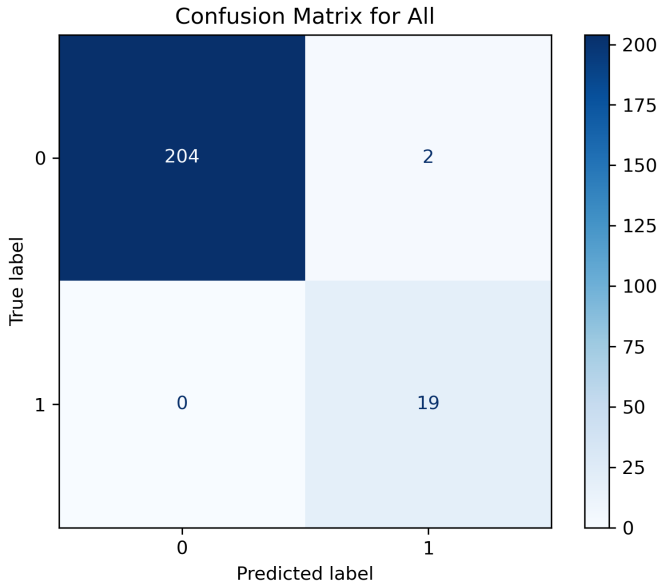

Figure 5: Confusion Matrixes for the category 'All', representing the category 'deaths in the context of an police intervention'.

The metrics measured evaluating the model with the test set have been the following:

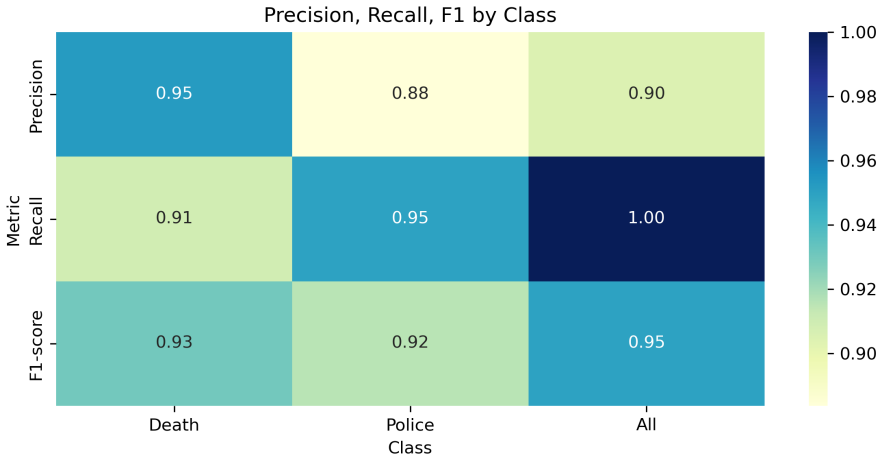

Figure 6: Metrics of the model. The category 'All' represents the category 'deaths in the context of an police intervention'

This model was then used to classify the titles and lead posts extracted in the previous step. Each text fragment receives a probability score between 0 and 1 for each of the three categories, indicating the likelihood that it pertains to that category.

The confusion matrix and overall evaluation metrics suggest that the model performs reasonably well. However, when applied to the real dataset, the results were less convincing. This issue will be addressed further in the 'Results' section.

3.2.6 Match text fragments with known categories

At first, an initial attempt was made to automatically extract information about death cases directly from the articles (LlamaIndex, Spacy entities). However, several issues arose. First, the model was not reliable enough to ensure that all top-ranked articles actually discussed deaths in the context of police interventions. Moreover, even when dealing with relevant articles, the entity extraction system struggled with subtle distinctions, such as differentiating between the name of a police officer and that of a victim. As a result, the decision was made not to rely on automatic data extraction. nstead, we chose to provide the list of articles that most likely discuss deaths in the context of police interventions that sorts from the model applicatoin.

Now, with a refined list of titles that most likely discuss deaths during police operations, the next step is to exclude articles that refer to events already included in our dataset. For this filtering process, the full content of the articles was scraped and added to the textual data. The text was then matched on the following categories:

Name
Location of death
Origin
Age
Date - Month & Year

If a name already appeared in the database, it served as a strong indicator that the article was referring to the death of that individual. In such cases, the article was assigned a score of '5', indicating a high likelihood that it described a case already recorded in the database. For other articles, a scoring system was used based on the presence of relevant information within the text. For example, if the article mentioned details such as the person's age or origin, and if it was published in the same year or month as the recorded death, it was assigned a score of 3 (fullfilled 3 categories), reflecting a moderate likelihood of referring to a known case. The final list could therefore be filtered as well on those matches, excluding the '5', and other high numbers. However, the articles are not removed from the dataset; rather, this serves as an additional tool to help readers filter the list more effectively.

3.2.7 Failed Scraping

During each scraping step, the system occasionally became overloaded, resulting in missing data (titles, lead posts, or article content). To address this, the failed links were stored and reused in a subsequent step using the same scraping code, which allowed us to recover the missing information and complete the dataset. For instance, the initial scraping of titles yielded 443 951 entries. After reprocessing the failed links, the total increased to 446 902 titles—an additional 2,951 entries. This repeated scraping process was necessary for all data types (titles, post leads, and content) to ensure full data retrieval.

4 Results

In the section 'Results', we are going to compare the two pipelines and discuss particularities encountered when analyzing the final List of Articles.

4.1 Evaluation of the two pipelines

4.1.1 Pipeline 1

Following the steps mentionned above, this leads us to a reduction of articles as following:

Start: 446902 → Reduction Keywords → 35552 → Reduction CH → 34324 → Model application

As it can be seen, the amount of articles remaining for the model to evaluate is still quite high. This presents especially a problem for the step of data scraping of the whole article, as this would be really time consuming for commercial computers. When having a look at the links leading towards the articles, it can be seen that still most of them are falling among the category 'monde' ('world'). The caution, not to erase Titles that might take place inside of Switzerland, will make a manual analysis of the final results way more complicated, as there remain too many articles. If possible when working with other newspaper, the second solution should therefore be choosen, where the location is reduced directly on particularities of the newspaper, such as in our case the links leading to them. For completing our database, we therefore used as well the outcome of Pipeline 2.

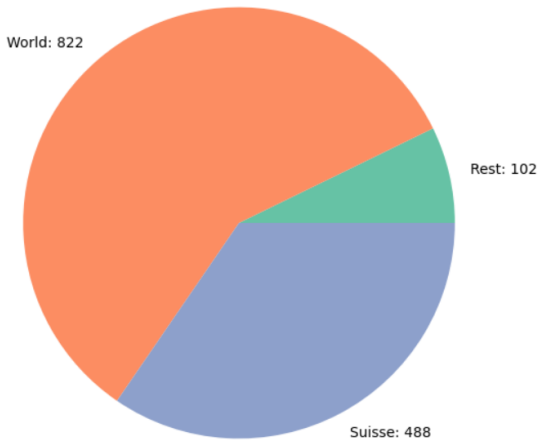

Figure 7: Repartition of articles between the category 'monde', 'suisse' and others of the newspaper 'Le Temps'

4.1.2 Pipeline 2

With this second pipeline, we obtained a reduction at the different steps as following:

Start: 446902 → Reduction CH → 89193 → Reduction Keywords → 10394 → Model application

In comparison with the first pipeline, this shows already a way more important reduction of articles. When running the model on the titles and lead posts, we obtain an excel with estimated probabilities of which the titles are talking about 'death in the context of a police intervention'. Furthermore, the titles are only in the category 'suisse' of the newspaper. This means however, that this pipeline is way less general than the first one, and will need more adaptations if used for other newspapers.

As conclusion, the results of the two pipelines were quite disillusioning. Although it arrived quite well to catch the presence of police or death / injury in the titles, it couldn't identify well the 'death during the police intervention'. The texts that were indicated to be of high probability to talk about 'death during the police intervention' were mostly just the texts talking about police as well as death. It isn't working well for distinguish if it is only a communiqué of the police, or if the police was on place in the case of the death. Nevertheless, the articles talking about deaths during police interventions could be found equally amongst the highest scoring articles.

4.2 List of promising titles

The output of the pipelines consisted of Excel lists containing the most promising article titles likely referring to deaths in the context of police interventions. While the model's performance appeared satisfactory based on its confusion matrix and evaluation metrics, applied to the real dataset it was less efficient. The model struggled particularly with nuanced distinctions. For example, it failed to differentiate between the death of a civilian and that of a police officer. It also confused contexts such as prisons (where police are not present) with police custody (where police is present). Additionally, the model frequently misclassified articles that mentioned both the police and a death, even when the death was unrelated to police intervention. For instance, a police press release about an unrelated murder investigation was often incorrectly flagged as a relevant case, despite not fitting the criteria.

text	Death	Police	All (lethal police violence	Predicted as 'Al	link	Content_	matche
Homme tué à Montreux: la police était en lég	0,95872	0,96622	0,976954818	TRUE	https://w	['Les policiers	2
Un policier genevois tire sur des voleurs en fu	0,958111	0,96812	0,976607144	TRUE	https://w	["Trois coups	2
La police interpelle un homme après une tent	0,96283	0,963326	0,976279199	TRUE	https://w	['Une tentativ	1
Un jeune policier tué par un proche dans la c	0,953813	0,970511	0,976208806	TRUE	https://w	['Un jeune po	2
Décès inexpliqué dans un poste de police gen	0,952847	0,971032	0,975998878	TRUE	https://w	["Un jeune hc	2
Militants et élus avant le procès de six policie	0,95715	0,969784	0,975851297	TRUE	https://w	['Une bander	5
La famille du détenu mort en prison porte pla	0,955624	0,969973	0,975816429	TRUE	https://w	['Me Nicolas M	1
Un policier d'élite vaudois tué par balle par u	0,964162	0,96162	0,97577697	TRUE	https://w	["Un policier v	1
Drame de Bex: le policier est acquitté Le sous	0,962194	0,965652	0,975712359	TRUE	https://w	['Le verdict n'	1
Drame conjugal à Genève: deux morts et un p	0,96366	0,965325	0,975536823	TRUE	https://w	['Une interver	1
Un second mineur en cause dans l'agression	0,962892	0,965035	0,975486815	TRUE	https://w	['Un autre mi	1
L'adolescente fribourgeoise a été retrouvée n	0,956992	0,968153	0,975484192	TRUE	https://w	['Un jeune Su	2
Une bagarre nocturne à Lausanne a fait un m	0,964732	0,963818	0,975255191	TRUE	https://w	['Durant la nu	2
Un détenu de Champ-Dollon meurt dans sa c	0,953949	0,969821	0,975123167	TRUE	https://w	['L'homme ét	1

Figure 8: List with most promising titles speaking about death in the context of a police intervention. The cases that were actually speaking about this are highlighted in green, the others in red.

All articles with a predicted probability higher than 96 % were reviewed. This threshold was chosen both due to time constraints and because the majority of articles classified with a score of 5 fell within this range. This cutoff resulted in a list of 173 articles, within which several new cases could be identified.

During the review of the articles, several ambiguous cases raised questions about whether they should be included. For instance, one case involved a girl who died of hunger and thirst after her mother was arrested, and the police failed to verify whether she had a child. This raised the question: does such a situation fall under the category of "death in the context of a police intervention"? (In this case, we decided to include it.) For such complex cases, decisions were made on an individual basis by myself, in consultation with my supervisor at Border Forensics. As a result, when presenting the final dataset, it is essential to also include the context surrounding each death. This ensures transparency and allows readers to understand the rationale behind the inclusion of each case.

4.3 New cases found

During the work on the datasets, a total of **19** new death cases could have been identified.**2** deaths needed to be added as they occured in the period during which the work was conducted. Furthermore, the work allowed to reveal the opportunities but also the limits of the natural language processing for such research work.

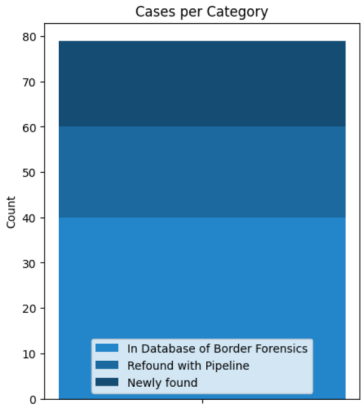

Figure 9: Distribution of cases by the different categories.

The cases of death during police interventions in the dataset could be classified at the end in three main categories: The ones, that were newly found during this semester projects. The ones, that were re-encountered with help of the Pipeline, where the title as well appeared in the highest percentage of Titles. And then the ones, that couldn't be found with help of the pipeline.

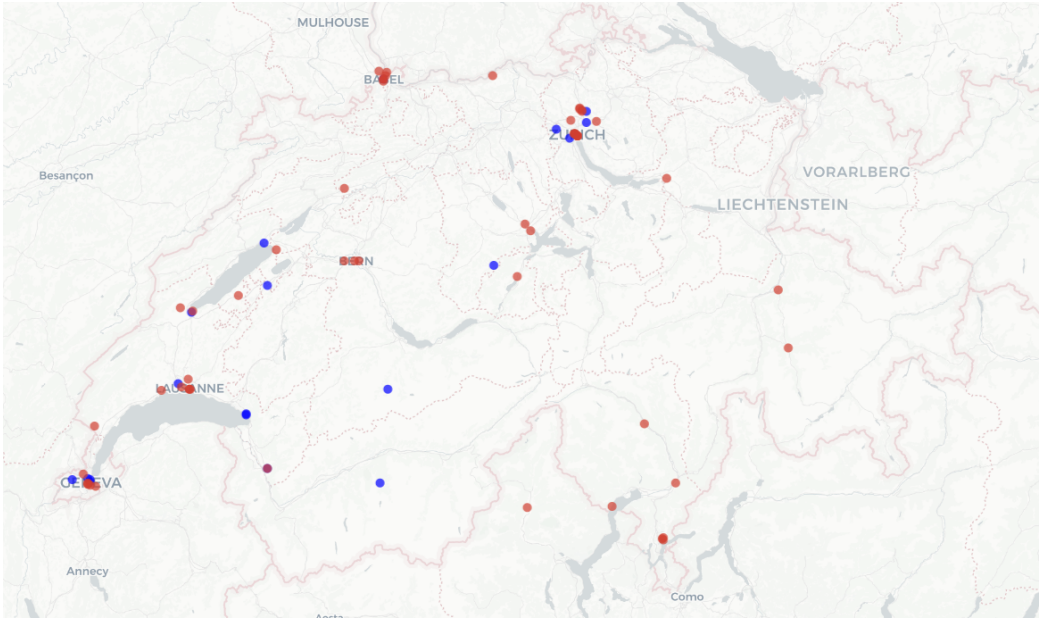

Figure 10: Map of the found an existing cases: Blue ware the points that were newly found during the project, red the ones that were already there before.

An analysis of the newly found or re-encountered cases reveals a clear trend toward the French-speaking part of Switzerland: the majority of these cases originate from that region.Given that the newspaper is French-language, its coverage is naturally focused on events in the French-speaking part of Switzerland. Of the 30 cases newly identified or re-encountered with the help of our model's output, only 13 came from non–French-speaking regions. By contrast, out of a total

of 41 undetected cases, only 8 were from Romandie (the French-speaking region). This suggests that a significant number of cases from the German- or Italian-speaking regions of Switzerland likely remain undiscovered, indicating that the dataset probably remains incomplete. This shows, that developing models for the other national languages remains an important next step, but was beyond the scope of what I could accomplish during my semester project. Furthermore, it should be noted that the *"True"* cases in the training set consisted of those that had already been identified and were already present in the dataset. Since the training set was in French, only cases that could be retrieved from French-language newspapers were included. As a result, the probability that these cases reappear in the final list is higher than for cases that did not occur in Switzerland or were not covered by French-language media.

4.4 Stepwise evaluation of code

Throughout the development of the two pipelines, I continuously assessed their performance to the extent possible. At each reduction step, the results were compared against the list of known cases to evaluate accuracy. In particular, I checked whether the cases of the database of Border Forensics that originally cited 'Le Temps' as a source were still being identified. If these reference cases remained present after the filtering step, the method used was considered appropriate.

4.5 Other data collection ways

Given the limitations observed in the pipeline's results, it becomes evident that this approach needs to be complemented by an additional method.

Through discussions with experts, crowdsourcing came up as one of the most promising alternatives. Building up a network of volunteers across Switzerland who monitor local media and report new cases of deaths in the context of police interventions, in their respective language regions. This grassroots method could potentially be even more effective, especially in capturing cases that automated tools or big newspaper might miss. A notable example is The Washington Post, which has taken the lead in documenting deaths resulting from police interventions in the United States [10]. Due to its wide public reach, new incidents are frequently reported to their newsroom, allowing for timely and comprehensive documentation. The message they posted on their website was the following:

> *To provide information about fatal police shootings since Jan. 1, 2015, send us an email at policeshootingsfeedback@washpost.com. The Post is also interested in obtaining photos of the deceased and original videos of fatal encounters with police.*

Such a method could be more effective in identifying additional cases that might have been overlooked until now.

5 Difficulties & Limits

5.1 Clear definition of cases to include

One of the major difficulties of the project was to define the cases to be included into the dataset. While at start the goal was to look for 'lethal police violence', it quickly became clear that this formulation is too vague. Whose definition of deadly police violence would we take? The one of the justice system, when only in case of recognition of the fault of a police officer the case would enter the database? Or our own judgement, but how would we define it?

Therefore we needed to break down the definition to include *any civilan person killed or deceased*

as a result of an interaction with a police or gendarmerie officer, whatever the circumstances that led to this intervention. This includes as well the cases, where the police wasn't even directly present in the moment of death, but where its actions led to the death. An example for it would be Lamin Fatty, who had an epileptic seizure that went unnoticed even though a surveillance camera was filming his cell in police custody. Or the little girl in Genève, that died of thirst and hunger at home after the arrestation of his mother.

5.2 Limits in the Code

5.2.1 Computational time

Due to the huge dataset, the computational time for the different steps was quite elevated. This complicates the use of such a pipeline for NGOs that don't have access to clusters allowing to reduce the computational time and to run different codes in parallel. A work that took already now a not negligible amount of time, would take even longer. For the project I started without using the Cluster of the DHLAB. However, in order to accelerate my work and to be able to finish the pipeline, towards the end I still needed to work with it. The estimation of the real time that it would take for a commercial computer without external help is therefore hard to estimate.

5.2.2 Human bias

As it has been shown by divers researcher the human behind a code induces as well a bias in the results coming out of the code [7]. It must be assumed that this is as well the case in this Pipeline. In order to counter this as much as possible, the choices made during the construction of the code has been explicited and explained in the report above. An example of this is the keyword reduction step: while we primarily relied on code (such as NLP similarity measures), we also incorporated a manually created list of words. This list, specified above for transparency, inevitably introduces some bias.

5.2.3 Efficiency of code vs human

Especially in the last step of the precise data extraction, the code worked not precise enough to distinguish between details of the articles, such as if the name belongs to of the police man, random person present or the person that died. Therefore it has been opted for leaving the final step to the human consulting the code, taking the articles that are most promising for giving new deaths.

However, the fact that a human needs to complete the last step allows a deeper understanding of the final database. Some of the cases are quite hard to assess, whether or not to take into account in the database. Those nuances can be decided and taken into account. An exammple for it is the case, where a police man is killed during a training session (not taken into account), or a police man killed during a police intervention (taken into account). Another one would be a person dying in police custody (taken into account), or a person dying in a prison (not taken into account).

5.3 Other limits

5.3.1 Estimation of veracity

Unlike France, in Switzerland there is no official registration yet of the deaths occuring in context of police interventions. A comparison of the datasets is therefore not possible, and it is difficult to estimate the veracity of the final dataset. Alternative techniques for estimating this veracity could be used, which however went beyond the scope of this project.

5.3.2 Language changes

The use of language changes in various ways. It varies already from one newspaper to another, from one country to another but as well over time.

Newspaper have different styles and political orientation of writing and using different words, influencing the performance of such a model. As an example, where in some newspaper a death is considered as a *malaise*, in others they are writing about *police killing*. The Pipeline was trained and worked with only 'Le Temps', meaning that the language used for training the model and reducing the sizes was only based on this newspaper. If other newspaper are added, for best precision, all the steps would needed to be repeated with the other newspaper and titles of the other newspaper would need to be included in the test-training dataset for the performance to be the same.

The vocabulary used for police in other countries are different languages. The model used was trained on Swiss Titles. Using it for other countries is therefore expected to perform less.

Language changes with the time. Expressions and words used to describe cases change with the time as well as from place to place. Its performance might as well decrease with time, as the language is changing ongoingly.

6 Further work

The semester work has infinite possibilities for further development. Some possibilites would be the addition of more Newspaper, adapting the model for other languages, finetuning the model on a larger test/train set, skipping the reduction steps and applying the model directly at start. Given the countless possibilities, I would like to go into detail only for some of the ideas.

6.1 Addition of impunity events

A further factor that would be of interest for Border Forensics is the legal status of the death in the cases where there was a trial: Was the case classified, how many time was an appeal made, did it pass to differnet courts (eg. from cantonal to national)... At the moment, the model isn't trained on juridicial events. In order to be able to extract this information, a fourth classification on 'jurisdiction' could be of high interest.

6.2 Languages & Newspapers

The model used at the moment is only done for french languages. Switzerland however containes three different main languages, as well as newspaper in all those three languages. In order to have a representative database, all of the three languages would need to be included. This however requires the adaptation of all the different steps for the different languages.

7 Considerations about Future Data Collection

The code is designed to be reusable in the coming years. Although computational time was a major challenge that motivated the development of these pipelines, focusing only on Le Temps over a relatively short period of time in the future means that the number of new titles added annually is likely to be limited. Consequently, the volume of new data to process each year will be considerably smaller.

However, when reusing this code in the future, all the limitations outlined above must be taken

into account. While the code should not be relied upon as the only source for identifying deaths in the context of police interventions, it can nevertheless serve as a useful tool to reduce and complete to a certain extent the work.

8 Critical reflection on working process

Reflecting on the development process of this semester project, there are several points of self-criticism I would like to address.

First, while I still consider consulting experts to be a crucial step, the development of the code should have started earlier to better integrate their input from the beginning. Having had the opportunity to speak to so many interessting actors in this field, I got at the start a bit lost with all the different possibilities and information, and advanced really slowly towards a concrete approach.

Additionally, motivated by the goal of creating reusable code, I initially worked only on my personal computer, which significantly slowed down the progress of the code. For future projects, I would recommend using resources such as computing clusters at an earlier stage to accelerate development. The evaluation of computational time could have been conducted later in the process.

Finally, due to the large dataset size, it was difficult to monitor the performance and correctness of each step throughout the process. This resulted in some errors only being discovered at a late stage. For future work, it will be essential to implement robust validation checks from the outset to ensure each part of the code functions correctly.

References

[1] Amnesty International. France: Des policiers au-dessus des lois. Rapport, Amnesty International, 2007. Consulté le 30 mai 2025.

[2] Augenauf. Augenauf – organisation suisse de défense des droits humains, 2025. Consulté le 30 mai 2025.

[3] Basta! Morts à la suite d'interventions policières, 2023. Consulté le 30 mai 2025.

[4] Border Forensics. Border forensics, 2025. Accessed on 30 May 2025.

[5] INQUEST. Deaths in police custody, 2025. Consulté le 30 mai 2025.

[6] Inspection générale de la Police nationale. Rapport annuel d'activité de l'igpn 2023. Rapport annuel, Ministère de l'Intérieur, France, novembre 2023. Consulté le 30 mai 2025.

[7] Johanna Johansen, Tore Pedersen, and Christian Johansen. Studying the transfer of biases from programmers to programs. *AI & SOCIETY: Knowledge, Culture and Communication*, 38(4):1659–1683, 2023.

[8] Le Temps. Archive — le temps. `https://www.letemps.ch/archive?srsltid=AfmBOorWUK_4Lg6M-ME91sNeCccWKehJhlSkW-mg7c4LhiOdwt3u5Zz1`, 2025. Archive search results page.

[9] Mapping Police Violence. Mapping police violence, 2025. Consulté le 30 mai 2025.

[10] The Washington Post. Fatal force: Police shootings database. `https://www.washingtonpost.com/graphics/investigations/police-shootings-database/`, 2022. Originally published May 30, 2015; updated December 5, 2022.

[11] United Nations Human Rights Council. Visit to switzerland – report of the working group of experts on people of african descent. Country Report A/HRC/51/54/Add.1, Office of the United Nations High Commissioner for Human Rights, 2022. Accessed on 30 May 2025.