



Jalon 1

8INF970 - Atelier pratique en cybersécurité II

Fehmi Jaafar      Samuel Desbiens

Justin Bossard      Mattéo Gouhier      Paul Mathé      Samuel Plet      Léo Raclet

25 janvier 2026

# Table des matières

|          |  |          |
|----------|--|----------|
| <b>1</b> | <b>Introduction</b>                          | <b>2</b> |
| <b>2</b> | <b>État de l'art</b>                         | <b>3</b> |
| <b>3</b> | <b>Notre solution</b>                        | <b>4</b> |
| 3.1      | Nom et logo . . . . .                        | 4        |
| 3.2      | Outils de développement . . . . .            | 4        |
| 3.3      | Outils de gestion . . . . .                  | 4        |
| <b>4</b> | <b>Choix des Données</b>                     | <b>5</b> |
| 4.1      | Stratégie de Sélection des Données . . . . . | 5        |
| 4.2      | Approche par Scraping . . . . .              | 5        |
| 4.3      | Jeux de Données Publics Reconnus . . . . .   | 5        |
| <b>5</b> | <b>Expérimentations possibles</b>            | <b>7</b> |
| <b>6</b> | <b>Conclusion</b>                            | <b>8</b> |
| <b>7</b> | <b>Références</b>                            | <b>9</b> |

## **Partie 1**

# **Introduction**

TODO

## **Partie 2**

### **État de l'art**

## **Partie 3**

# **Notre solution**

### **3.1 Nom et logo**

### **3.2 Outils de développement**

### **3.3 Outils de gestion**

Pour la gestion de projet, nous communiquons via WhatsApp. Le code et les sources des rapports sont partagées sur [GitHub](#).

## Partie 4

# Choix des Données

### 4.1 Stratégie de Sélection des Données

L'entraînement d'une IA à pour but de donner de nombreux exemples de ce que l'on cherche afin quelle apprenne à le reconnaître. Dans notre cas, les fake news et les informations vérifiées. Pour permettre à une IA de fonctionner aux mieux dans un cas précis, il faut donc l'entraîner à l'aide d'un jeu de données divers et adapté à son utilisation. Nous avons identifié deux approches pour constituer une base de données la plus adaptées : le scraping de sources existantes et l'utilisation d'un jeu de données publics.

### 4.2 Approche par Scraping

Le scraping est une méthode qui consiste à parcourir automatiquement des pages web et en extraire ses données. Dans notre cas, deux sources pourraient s'appliquer :

#### Sources de Contenu Vérifié et Fiable :

- Regrouper plusieurs sites d'informations reconnus : Ces sources constituent une base de données d'articles vérifiés, permettant d'établir des références que l'on sait justes pour un modèle d'apprentissage. Permettant d'avoir un grand nombre de données tout en étant certains de l'exactitude des informations. Par exemple, c'est ce que fait la plateforme de vérification Vera, qui regroupe environ 300 sites de confiance.

#### Sources de Contenu Trompeur :

- Site regroupant des fake News : Afin d'avoir des informations vérifiées, nous pourrions directement aller sur une plateformes qui références les fake news ou les informations qui peuvent prêter à la confusion, le but de ces plateformes/rubriques étant de valider la véracité ou non des informations. Par exemple, la rubrique Vrai/Faux de FranceInfo. La plateforme de FranceInfo maintient une rubrique spécialisée qui analyse les affirmations fausses ou trompeuses. Cette source est très intéressante car elle fournit l'identification des fausses informations, mais aussi une analyse et des explications détaillées des raisons pour lesquelles ces affirmations sont erronées, tout en étant rattachée à un organisme fiable.

### 4.3 Jeux de Données Publics Reconnus

Au-delà du scraping, permettant de se créer un jeu de données, nous pourrions aussi nous appuyer sur des jeux de données publics existants, largement utilisés, testés et reconnus :

- **Webz.io** : Une base de données indexant des millions d'articles web en anglais, avec des métadonnées riches, permettant une analyse du contexte.

- **ISOT Fake News Dataset** : Un jeu de données spécialisé contenant plusieurs milliers d’articles validés comme vrais ou faux, avec des étiquetages de très bonne qualité.
- **FakeNewsNet** : Un ensemble de données complet incluant non seulement le contenu des articles.

Ces jeux de données permettent d’avoir des articles vérifiés ainsi que des informations supplémentaires permettant d’avoir du contexte ou de distinguer certains articles, qui pourront permettre à l’IA de gagner en précision.

## Partie 5

# Expérimentations possibles

## **Partie 6**

### **Conclusion**

TODO

## Partie 7

# Références