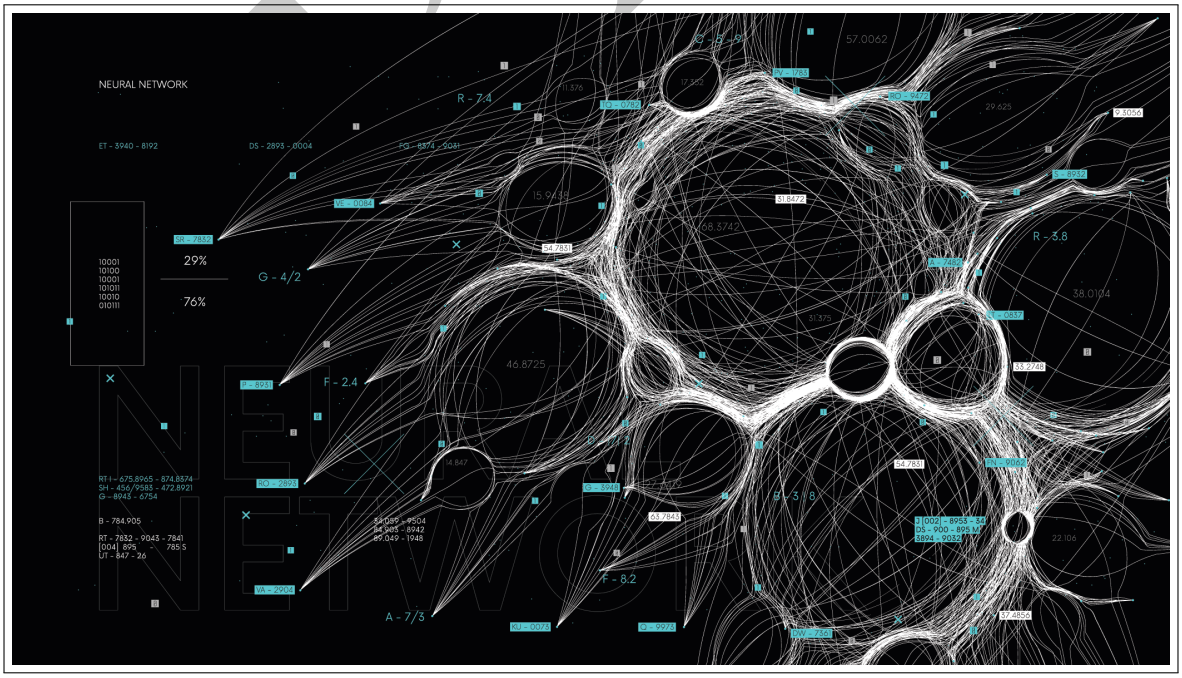
Institut National des Sciences Appliquées
Université de Rennes

Étude d'Attaques par Inférence d'Appartenance (MIA)

Une première approche avec le concours Snake Strikes Back

Auteurs :
Thomas Aubin
Selyan Da Silva
Moussa Ouassou
Émile Peltier

Responsable de projet :
Cédric Eichler,
Enseignant-Chercheur en Informatique
Relecteurs :
Prénom Nom



Résumé en quelques lignes du projet

Mots-clés : mot1, mot2, mot3, mot4, mot5

16 novembre 2024

Résumé

Ceci est un exemple d'encadré. Il sert à mettre en évidence des parties importantes du rapport

Table des matières

Table des figures

Liste des tableaux

Liste des Équations

Table des éléments de code

Introduction

Bien que le projet ait pour coeur la participation à la compétition, celui-ci a nécessité un important travail de montée en compétences et de documentation en Machine Learning pour l'ensemble du groupe, ce domaine n'étant que peu abordé à ce stade de la formation. C'est pourquoi la partie opérationnelle et technique du projet est précédée d'une part d'un court travail de bibliographie ayant pour visée la synthèse des connaissances mathématiques et algorithmiques indispensables à la participation au concours, et d'autre part par une présentation des tenants et aboutissants du concours, laquelle prend soin d'expliquer le plus finement possible les données sur lesquelles nous nous entraînons ainsi que le modèle attaqué.

Première partie

Notions d'*Adversarial Machine Learning*

Chapitre I

Introduction aux concepts utiles d'Intelligence Artificielle

I.1 Section 1.1

I.1.1 Sous-section 1.1.1

I.1.1.1 Sous-sous-section 1.1.1.1

$$a + b = c \tag{I.1}$$

FIGURE I.1 – Exemple de figure

I.1.1.2 Sous-sous-section 1.1.1.2

```
1 print("This line will be printed.")
2 print("Another line to print."
```

Listing I.1 – Un code Python

I.1.1.3 Sous-sous-section 1.1.1.3

I.1.2 Sous-section 1.1.2

I.1.2.1 Sous-sous-section 1.1.2.1

I.1.2.2 Sous-sous-section 1.1.2.2

I.1.2.3 Sous-sous-section 1.1.2.3

I.1.3 Sous-section 1.1.3

I.1.3.1 Sous-sous-section 1.1.3.1

I.1.3.2 Sous-sous-section 1.1.3.2

I.1.3.3 Sous-sous-section 1.1.3.3

I.2 Section 1.2

I.2.1 Sous-section 1.2.1

I.2.1.1 Sous-sous-section 1.2.1.1

I.2.1.2 Sous-sous-section 1.2.1.2

I.2.1.3 Sous-sous-section 1.2.1.3

I.2.2 Sous-section 1.2.2

I.2.2.1 Sous-sous-section 1.2.2.1

I.2.2.2 Sous-sous-section 1.2.2.2

I.2.2.3 Sous-sous-section 1.2.2.3

I.2.3 Sous-section 1.2.3

I.2.3.1 Sous-sous-section 1.2.3.1

I.2.3.2 Sous-sous-section 1.2.3.2

I.2.3.3 Sous-sous-section 1.2.3.3

I.3 Réseaux de neurones et Deep Learning

I.3.1 Sous-section 1.3.1

I.3.1.1 Sous-sous-section 1.3.1.1

I.3.1.2 Sous-sous-section 1.3.1.2

I.3.1.3 Sous-sous-section 1.3.1.3

I.3.2 Sous-section 1.3.2

I.3.2.1 Sous-sous-section 1.3.2.1

I.3.2.2 Sous-sous-section 1.3.2.2

I.3.2.3 Sous-sous-section 1.3.2.3

I.3.3 Un modèle à deux réseaux : le *Generative Adversarial Network* (GAN)

I.3.3.1 Sous-sous-section 1.3.3.1

I.3.3.2 Sous-sous-section 1.3.3.2

I.3.3.3 Sous-sous-section 1.3.3.3

Chapitre II

Chapitre II

II.1 Section II.1

II.1.1 Sous-section II.1.1

II.1.1.1 Sous-sous-section II.1.1.1

II.1.1.2 Sous-sous-section II.1.1.2

II.1.1.3 Sous-sous-section II.1.1.3

II.1.2 Sous-section II.1.2

II.1.2.1 Sous-sous-section II.1.2.1

II.1.2.2 Sous-sous-section II.1.2.2

II.1.2.3 Sous-sous-section II.1.2.3

II.1.3 Sous-section II.1.3

II.1.3.1 Sous-sous-section II.1.3.1

II.1.3.2 Sous-sous-section II.1.3.2

II.1.3.3 Sous-sous-section II.1.3.3

II.2 Section II.2

II.2.1 Sous-section II.2.1

II.2.1.1 Sous-sous-section II.2.1.1

II.2.1.2 Sous-sous-section II.2.1.2

II.2.1.3 Sous-sous-section II.2.1.3

II.2.2 Sous-section II.2.2

II.2.2.1 Sous-sous-section II.2.2.1

II.2.2.2 Sous-sous-section II.2.2.2

II.2.2.3 Sous-sous-section II.2.2.3

II.2.3 Sous-section II.2.3

II.2.3.1 Sous-sous-section II.2.3.1

II.2.3.2 Sous-sous-section II.2.3.2

II.2.3.3 Sous-sous-section II.2.3.3

II.3 Section II.3

II.3.1 Sous-section II.3.1

II.3.1.1 Sous-sous-section II.3.1.1

II.3.1.2 Sous-sous-section II.3.1.2

II.3.1.3 Sous-sous-section II.3.1.3

II.3.2 Sous-section II.3.2

II.3.2.1 Sous-sous-section II.3.2.1

II.3.2.2 Sous-sous-section II.3.2.2

II.3.2.3 Sous-sous-section II.3.2.3

II.3.3 Sous-section II.3.3

II.3.3.1 Sous-sous-section II.3.3.1

II.3.3.2 Sous-sous-section II.3.3.2

II.3.3.3 Sous-sous-section II.3.3.3

Chapitre III

Attaques par Inférences d'Appartenance : contextualisation du projet

III.1 Section III.1

III.1.1 Sous-section III.1.1

III.1.1.1 Sous-sous-section III.1.1.1

III.1.1.2 Sous-sous-section III.1.1.2

III.1.1.3 Sous-sous-section III.1.1.3

III.1.2 Sous-section III.1.2

III.1.2.1 Sous-sous-section III.1.2.1

III.1.2.2 Sous-sous-section III.1.2.2

III.1.2.3 Sous-sous-section III.1.2.3

III.1.3 Sous-section III.1.3

III.1.3.1 Sous-sous-section III.1.3.1

III.1.3.2 Sous-sous-section III.1.3.2

III.1.3.3 Sous-sous-section III.1.3.3

III.2 Section III.2

III.2.1 Sous-section III.2.1

III.2.1.1 Sous-sous-section III.2.1.1

III.2.1.2 Sous-sous-section III.2.1.2

III.2.1.3 Sous-sous-section III.2.1.3

III.2.2 Sous-section III.2.2

III.2.2.1 Sous-sous-section III.2.2.1

III.2.2.2 Sous-sous-section III.2.2.2

III.2.2.3 Sous-sous-section III.2.2.3

III.2.3 Sous-section III.2.3

III.2.3.1 Sous-sous-section III.2.3.1

III.2.3.2 Sous-sous-section III.2.3.2

III.2.3.3 Sous-sous-section III.2.3.3

III.3 Section III.3

III.3.1 Sous-section III.3.1

III.3.1.1 Sous-sous-section III.3.1.1

III.3.1.2 Sous-sous-section III.3.1.2

III.3.1.3 Sous-sous-section III.3.1.3

III.3.2 Sous-section III.3.2

III.3.2.1 Sous-sous-section III.3.2.1

III.3.2.2 Sous-sous-section III.3.2.2

III.3.2.3 Sous-sous-section III.3.2.3

III.3.3 Sous-section III.3.3

III.3.3.1 Sous-sous-section III.3.3.1

III.3.3.2 Sous-sous-section III.3.3.2

III.3.3.3 Sous-sous-section III.3.3.3

Deuxième partie

Le concours *Snake Strikes Back* : position du problème

Chapitre IV

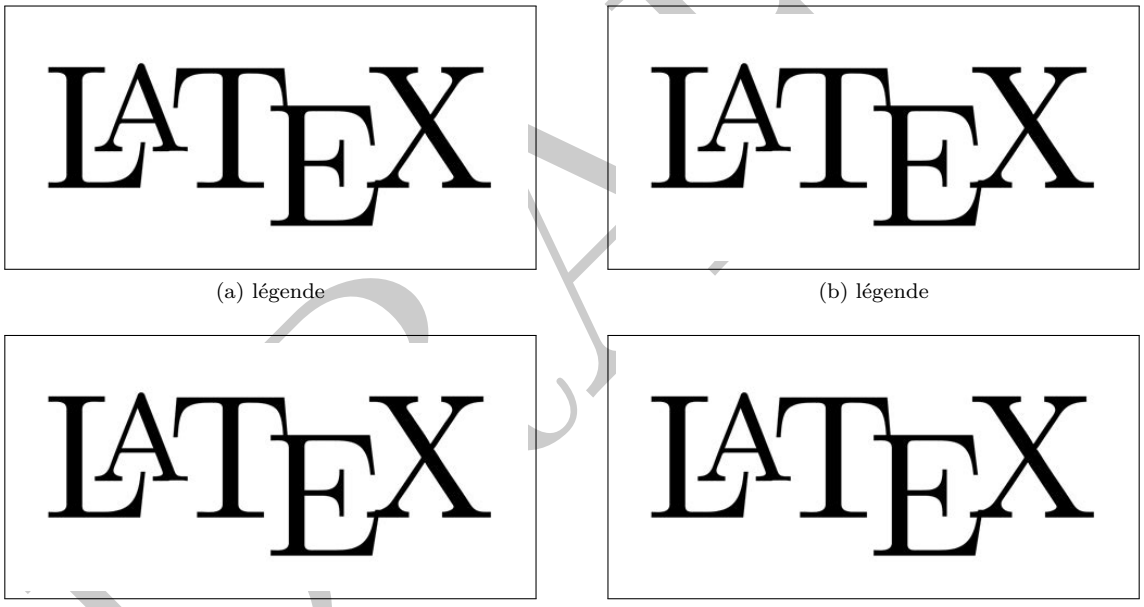
Contexte et enjeux de la compétition

IV.1 Section IV.1

IV.1.1 Sous-section IV.1.1

IV.1.1.1 Sous-sous-section IV.1.1.1

IV.1.1.2 Sous-sous-section IV.1.1.2



(a) légende (b) légende

(c) légende (d) légende

FIGURE IV.1 – Exemple de figure avec plusieurs images

IV.1.1.3 Sous-sous-section IV.1.1.3

IV.1.2 Sous-section IV.1.2

IV.1.2.1 Sous-sous-section IV.1.2.1

IV.1.2.2 Sous-sous-section IV.1.2.2

IV.1.3 Sous-section IV.1.3

IV.1.3.1 Sous-sous-section IV.1.3.1

IV.1.3.2 Sous-sous-section IV.1.3.2

IV.1.3.3 Sous-sous-section IV.1.3.3

IV.2 Section IV.2

IV.2.1 Sous-section IV.2.1

IV.2.1.1 Sous-sous-section IV.2.1.1

IV.2.1.2 Sous-sous-section IV.2.1.2

IV.2.1.3 Sous-sous-section IV.2.1.3

IV.2.2 Sous-section IV.2.2

IV.2.2.1 Sous-sous-section IV.2.2.1

IV.2.2.2 Sous-sous-section IV.2.2.2

IV.2.2.3 Sous-sous-section IV.2.2.3

IV.2.3 Sous-section IV.2.3

IV.2.3.1 Sous-sous-section IV.2.3.1

IV.2.3.2 Sous-sous-section IV.2.3.2

IV.2.3.3 Sous-sous-section IV.2.3.3

IV.3 Section IV.3

IV.3.1 Sous-section IV.3.1

IV.3.1.1 Sous-sous-section IV.3.1.1

IV.3.1.2 Sous-sous-section IV.3.1.2

IV.3.1.3 Sous-sous-section IV.3.1.3

IV.3.2 Sous-section IV.3.2

IV.3.2.1 Sous-sous-section IV.3.2.1

IV.3.2.2 Sous-sous-section IV.3.2.2

IV.3.2.3 Sous-sous-section IV.3.2.3

IV.3.3 Sous-section IV.3.3

IV.3.3.1 Sous-sous-section IV.3.3.1

IV.3.3.2 Sous-sous-section IV.3.3.2

IV.3.3.3 Sous-sous-section IV.3.3.3

Chapitre V

Parcours des ressources fournies

V.1 Section

V.1.1 Sous-section

V.1.1.1 Sous-sous-section

V.1.1.2 Sous-sous-section

V.1.1.3 Sous-sous-section

V.1.2 Sous-section

V.1.2.1 Sous-sous-section

V.1.2.2 Sous-sous-section

V.1.2.3 Sous-sous-section

V.1.3 Sous-section

V.1.3.1 Sous-sous-section

V.1.3.2 Sous-sous-section

V.1.3.3 Sous-sous-section

V.2 Section

V.2.1 Sous-section

V.2.1.1 Sous-sous-section

V.2.1.2 Sous-sous-section

V.2.1.3 Sous-sous-section

V.2.2 Sous-section

V.2.2.1 Sous-sous-section

V.2.2.2 Sous-sous-section

V.2.2.3 Sous-sous-section

V.2.3 Sous-section

V.2.3.1 Sous-sous-section

V.2.3.2 Sous-sous-section

V.2.3.3 Sous-sous-section

V.3 Section

V.3.1 Sous-section

V.3.1.1 Sous-sous-section

V.3.1.2 Sous-sous-section

V.3.1.3 Sous-sous-section

V.3.2 Sous-section

V.3.2.1 Sous-sous-section

V.3.2.2 Sous-sous-section

V.3.2.3 Sous-sous-section

V.3.3 Sous-section

V.3.3.1 Sous-sous-section

V.3.3.2 Sous-sous-section

V.3.3.3 Sous-sous-section

Chapitre VI

DoppelGANger : un générateur de séries temporelles puissant ... mais attaquable

VI.1 Section

VI.1.1 Sous-section

VI.1.1.1 Sous-sous-section

VI.1.1.2 Sous-sous-section

VI.1.1.3 Sous-sous-section

VI.1.2 Sous-section

VI.1.2.1 Sous-sous-section

VI.1.2.2 Sous-sous-section

VI.1.2.3 Sous-sous-section

VI.1.3 Sous-section

VI.1.3.1 Sous-sous-section

VI.1.3.2 Sous-sous-section

VI.1.3.3 Sous-sous-section

VI.2 Section

VI.2.1 Sous-section

VI.2.1.1 Sous-sous-section

VI.2.1.2 Sous-sous-section

VI.2.1.3 Sous-sous-section

VI.2.2 Sous-section

VI.2.2.1 Sous-sous-section

VI.2.2.2 Sous-sous-section

VI.2.2.3 Sous-sous-section

VI.2.3 Sous-section

VI.2.3.1 Sous-sous-section

VI.2.3.2 Sous-sous-section

VI.2.3.3 Sous-sous-section

VI.3 Les hyperparamètres du modèles

VI.3.1 Sous-section

VI.3.1.1 Sous-sous-section

VI.3.1.2 Sous-sous-section

VI.3.1.3 Sous-sous-section

VI.3.2 Sous-section

VI.3.2.1 Sous-sous-section

VI.3.2.2 Sous-sous-section

VI.3.2.3 Sous-sous-section

VI.3.3 Sous-section

VI.3.3.1 Sous-sous-section

VI.3.3.2 Sous-sous-section

VI.3.3.3 Sous-sous-section

Troisième partie

Attaque d'un modèle de Machine Learning : méthodologie et raisonnements

Chapitre VII

Création de *Shadow Models*)pour reproduire le comportement étudié

VII.1 Critères déterminants dans la construction du modèle

VII.1.1 Overfitting

VII.1.1.1 Sous-sous-section

VII.1.1.2 Sous-sous-section

VII.1.1.3 Choix de la métrique

VII.1.2 Sous-section

VII.1.2.1 Structure

VII.1.2.2 Type

VII.1.3 Sous-section

VII.1.3.1 Sous-sous-section

VII.1.3.2 Sous-sous-section

VII.1.3.3 Sous-sous-section

VII.2 Section

VII.2.1 Sous-section

VII.2.1.1 Sous-sous-section

VII.2.1.2 Sous-sous-section

VII.2.1.3 Sous-sous-section

VII.2.2 Sous-section

VII.2.2.1 Sous-sous-section

VII.2.2.2 Sous-sous-section

VII.2.2.3 Sous-sous-section

VII.2.3 Sous-section

VII.2.3.1 Sous-sous-section

VII.2.3.2 Sous-sous-section

VII.2.3.3 Sous-sous-section

VII.3 Section

VII.3.1 Sous-section

VII.3.1.1 Sous-sous-section

VII.3.1.2 Sous-sous-section

VII.3.1.3 Sous-sous-section

VII.3.2 Sous-section

VII.3.2.1 Sous-sous-section

VII.3.2.2 Sous-sous-section

VII.3.2.3 Sous-sous-section

VII.3.3 Sous-section

VII.3.3.1 Sous-sous-section

VII.3.3.2 Sous-sous-section

VII.3.3.3 Sous-sous-section

Chapitre VIII

Chapitre

VIII.1 Section

VIII.1.1 Sous-section

VIII.1.1.1 Sous-sous-section

VIII.1.1.2 Sous-sous-section

$$a + b = c \tag{VIII.1}$$

VIII.1.1.3 Sous-sous-section

VIII.1.2 Sous-section

VIII.1.2.1 Sous-sous-section

VIII.1.2.2 Sous-sous-section

VIII.1.2.3 Sous-sous-section

VIII.1.3 Sous-section

VIII.1.3.1 Sous-sous-section

VIII.1.3.2 Sous-sous-section

VIII.1.3.3 Sous-sous-section

VIII.2 Section

VIII.2.1 Sous-section

VIII.2.1.1 Sous-sous-section

VIII.2.1.2 Sous-sous-section

VIII.2.1.3 Sous-sous-section

VIII.2.2 Sous-section

VIII.2.2.1 Sous-sous-section

VIII.2.2.2 Sous-sous-section

VIII.2.2.3 Sous-sous-section

VIII.2.3 Sous-section

VIII.2.3.1 Sous-sous-section

VIII.2.3.2 Sous-sous-section

VIII.2.3.3 Sous-sous-section

VIII.3 Section

VIII.3.1 Sous-section

VIII.3.1.1 Sous-sous-section

VIII.3.1.2 Sous-sous-section

VIII.3.1.3 Sous-sous-section

VIII.3.2 Sous-section

VIII.3.2.1 Sous-sous-section

VIII.3.2.2 Sous-sous-section

VIII.3.2.3 Sous-sous-section

VIII.3.3 Sous-section

VIII.3.3.1 Sous-sous-section

VIII.3.3.2 Sous-sous-section

VIII.3.3.3 Sous-sous-section

Chapitre IX

Synthèse des résultats

IX.1 Tâche 1

IX.1.1 Sous-section

IX.1.1.1 Sous-sous-section

IX.1.1.2 Sous-sous-section

IX.1.1.3 Sous-sous-section

IX.1.2 Sous-section

IX.1.2.1 Sous-sous-section

IX.1.2.2 Sous-sous-section

IX.1.2.3 Sous-sous-section

IX.1.3 Sous-section

IX.1.3.1 Sous-sous-section

IX.1.3.2 Sous-sous-section

IX.1.3.3 Sous-sous-section

IX.2 Tâche 2

IX.2.1 Sous-section

IX.2.1.1 Sous-sous-section

IX.2.1.2 Sous-sous-section

IX.2.1.3 Sous-sous-section

IX.2.2 Sous-section

IX.2.2.1 Sous-sous-section

IX.2.2.2 Sous-sous-section

IX.2.2.3 Sous-sous-section

IX.2.3 Sous-section

IX.2.3.1 Sous-sous-section

IX.2.3.2 Sous-sous-section

IX.2.3.3 Sous-sous-section

IX.3 Tâche 3

IX.3.1 Sous-section

IX.3.1.1 Sous-sous-section

IX.3.1.2 Sous-sous-section

IX.3.1.3 Sous-sous-section

IX.3.2 Sous-section

IX.3.2.1 Sous-sous-section

IX.3.2.2 Sous-sous-section

IX.3.2.3 Sous-sous-section

IX.3.3 Sous-section

IX.3.3.1 Sous-sous-section

IX.3.3.2 Sous-sous-section

IX.3.3.3 Sous-sous-section

IX.4 Tâche 4

IX.4.1 Sous-section

IX.4.1.1 Sous-sous-section

IX.4.1.2 Sous-sous-section

IX.4.1.3 Sous-sous-section

IX.4.2 Sous-section

IX.4.2.1 Sous-sous-section

IX.4.2.2 Sous-sous-section

IX.4.2.3 Sous-sous-section

IX.4.3 Sous-section

IX.4.3.1 Sous-sous-section

IX.4.3.2 Sous-sous-section

IX.4.3.3 Sous-sous-section

Conclusion

Quatrième partie

Annexes

Annexe 1 : Programmes conçus par l'équipe

Annexe 2 : Retour d'expérience et chronologie du projet

Annexe 3 : Framework utilisé

Bibliographie

[1] Tristan Allard et Mathias Bernard. « Snakes Strikes Back ». In : (oct. 2024).

[2] Author. *Membership inference attacks from first principles*. How published. Some note. Month Year. url : https://www.youtube.com/watch?v=1CNxfhMlk-A.

[3] Author. *Title*. How published. Some note. Month Year. url : URL.

[4] Zinan Lin et al. « Using GANs for Sharing Networked Time Series Data : Challenges, Initial Promise, and Open Questions ». In : (jan. 2021). Présentation du modèle DoppelGANger. url : https://arxiv.org/abs/1909.13403.

[5] Author Name. « Title of the Article ». In : *Journal Name* Volume Number.Issue Number (Month Year). Additional Notes, Start-End Pages. issn : ISSN Number. doi : DOI. url : URL.

[6] Author Name. *Title of the Book*. Edition (e.g., 2nd). T. Volume Number. Series Title. Additional Notes. Publisher Address : Publisher Name, Month Year of Publication. isbn : ISBN Number.

[7] Author Name. *Title of the Booklet*. How It Was Published. Additional Notes. Publisher Address, Month Year.

[8] Author Name. « Title of the Chapter ». In : *Title of the Book*. Sous la dir. d'Editor(s) Name. Edition (e.g., 2nd). T. Volume Number. Series Title Issue Number. Additional Notes. Publisher Address : Publisher Name, Month Year of Publication. Chap. Chapter Number, Start-End Pages.

[9] Author Name. « Title of the Paper ». In : *Title of the Conference Proceedings*. Sous la dir. d'Editor(s) Name. T. Volume Number. Series Title Issue Number. Additional Notes. Sponsoring Organization. Publisher Address : Publisher Name, Month Year, Start-End Pages.

[10] Author Name. *Title of the Report*. Type of Report Report Number. Additional Notes. Institution Address : Institution Name, Month Year. url : URL.

[11] Author Name. « Title of the Thesis ». Additional Notes. Type of Thesis (e.g., PhD). School Address : University Name, Month Year of Completion. url : URL.

[12] Author Name. « Title of the Thesis ». Additional Notes. Type of Thesis (e.g., Master's). School Address : University Name, Month Year of Completion. url : URL.

[13] Author Name. « Title of the Work ». Note (e.g., forthcoming, in preparation). Month Year.

[14] Reza Shokri. *Membership Inference Attacks against Machine Learning Models*. En ligne. Vidéo de vulgarisation du papier du même nom. Mai 2017. url : https://www.youtube.com/watch?v=rDm1n2gceJY&t=53s.

[15] Reza Shokri et al. « Membership Inference Attacks Against Machine Learning Models ». In : ().