# Proposition de support technique

## Objectifs

Le cadre proposé ici est volontairement large, le but n'est pas de tout implémenter mais de proposer du code pour alimenter une discussion d'environ 45 minutes

## Ressources

- L'archive data.zip. Le mot de passe pour l'ouvrir est : **infected**. Elle contient deux dossiers
  - pdfs : des rapports de cybersécurité
  - stix_bundles : des *bundles Stix*, qui sont les informations qui modélisent les rapports au format Stix, défini par Oasis (ref: https://oasis-open.github.io/cti-documentation/stix/walkthrough)
- D'autres rapport de cybersécurité disponibles en ligne : https://github.com/CyberMonitor/APT_CyberCriminal_Campagin_Collections/tree/master
- Le jupyter notebook "Prise en main.ipynb"
  - Attention, il y a trois paquets requis : "pdfplumber", "gliner" et "spacy"

## Pistes de travail

- Comment réconcilier les informations d'un *bundle stix* avec le fichier pdf associé ?
- Quelle modélisation utiliserais-tu pour extraire les *Malware* d'un fichier pdf ? Comment mesurerais-tu la pertinence de ce modèle ?
- On souhaite visualiser la pertinence d'une extraction *via* une interface web : quel outil utiliserais-tu ? Comment le déploierais-tu ?
- Comment ingérerais-tu automatiquement les nouveaux rapports du dépôt Github ?
- Tout ce que les ressources peuvent t'inspirer, en lien avec les missions identifiées dans la fiche de poste !