

# Offre de stage :

## Automates pour la vérification de modèles issus de l'apprentissage

---

**Encadrant:** Nathanaël Fijalkow (CNRS and the Alan Turing Institute of data science and artificial intelligence)

**Localisation:** LaBRI (Bordeaux)

**Sujet du stage:** Ce stage propose d'étudier différents algorithmes pour la vérification de modèles issus de l'apprentissage, dont en particulier les réseaux de neurones. Plusieurs approches sont proposées, basées soit sur des abstractions d'automates ou sur la synthèse d'invariants. On adoptera un point de vue plus théorique ou plus pratique selon le déroulement du stage et l'intérêt du stagiaire.

**Thèmes:** Théorie des automates, Apprentissage, Systèmes Dynamiques, Théorie des jeux, Vérification, Interprétation Abstraite.

---

Les réseaux de neurones et autres modèles en apprentissage ont récemment démontré leur utilité dans un grand nombre d'applications. Mais pour être déployé dans un scénario critique il faut des garanties : le modèle est-il précis, les résultats sont-ils robustes ?

Un exemple très classique montre un réseau de neurone dont l'objectif est de déterminer quel animal est sur une image, qui est trompé par l'ajout de bruit dans l'image : un panda devient un singe !

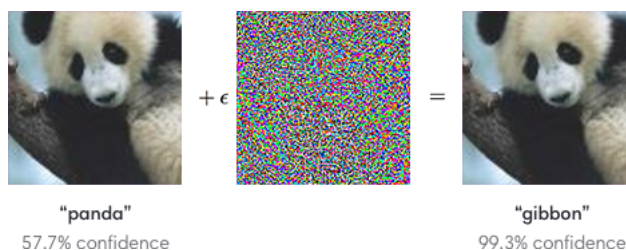


Figure 1: Issu de `openai.com`

La vérification des modèles d'apprentissage est un domaine de recherche jeune et dynamique, inspiré par la théorie des automates, la vérification de programmes, la synthèse d'invariants, et les systèmes dynamiques.

### Objectifs du stage

L'objectif du stage est d'étudier et de construire des algorithmes pour vérifier des propriétés simples sur des réseaux de neurones. La première étape sera de définir et comparer différentes notions de corrections et de robustesse. Les résultats du stage peuvent être soit théoriques soit pratiques par la mise en œuvre d'une implémentation.

Une bourse de thèse est disponible pour commencer en septembre 2019.