



*english follows*

## **Stage en intelligence artificielle - développement d'un algorithme de reconnaissance visuelle pour identifier des espèces de papillons à partir de photographies**

Durée: 8 mois ( avec possibilité de prolongation de 4 mois)

Supervision: David Rolnick (Université McGill, Mila) et Maxim Larrivée (Insectarium de Montréal, IRBV)

Lieux: Télétravail, accès aux serveurs à distance

Rémunération: 30 000 \$ (si la demande de bourse Mitacs est acceptée)

Période d'affichage: 5 janvier 2022 - 21 janvier 2022

Contexte:

[eButterfly](#) est un programme de science participative lancé en 2012 qui vise à documenter la biogéographie des lépidoptères. Initialement limité au Canada et aux États-Unis, le programme a été étendu en 2021 au Mexique, à l'Amérique centrale et aux Caraïbes. À ce jour, près de 500 000 observations ont été soumises, représentant 1 047 espèces. En 2022, l'objectif est d'étendre le programme au monde entier.

Depuis 2020, l'équipe de eButterfly collabore avec des chercheurs en intelligence artificielle (Mila, Centrale Supélec) pour développer un algorithme de reconnaissance visuelle pour assister l'identification des observations soumises. Un modèle a été entraîné en utilisant les images disponibles sur eButterfly et sur d'autres plateformes comme GBIF, et en intégrant des métadonnées relatives à l'endroit et au moment où les images ont été prises. Le modèle a été baptisé [eButterfAI](#).

Description du stage:

Le projet a pour objectif général le développement d'un outil permettant d'identifier automatiquement des papillons de jour de tous les continents à partir d'images de spécimens



vivants et naturalisés (épinglés) soumises à eButterfly. Cet outil permettra aux participants et aux experts de faciliter et accélérer l'identification et la validation des observations soumises.

La première étape consistera en l'agrégation d'images de différentes sources pour obtenir un jeu d'images dans lequel plus de 18 000 espèces (classes) sont représentées. Les sources d'images incluront eButterfly, GBIF, Leps by Fieldguide, BugGuide et iNaturalist. Un modèle sera ensuite entraîné en utilisant ces images et des "a priori" notamment la biologie (phénologie, répartition géographique) des espèces. Le modèle devra être hiérarchique et donner un résultat pour l'identification à la famille, au genre et à l'espèce.

Il est attendu que les papillons photographiés soient identifiés au niveau de la famille avec un haut taux de succès (>95%). Au niveau du genre et de l'espèce, le taux de succès devrait être plus grand ou égal à 80% dans la majorité des cas. Un taux plus bas serait initialement acceptable pour les classes pour lesquelles peu d'images sont disponibles.

#### Profil recherché:

- Étudiant.e au doctorat
- Expérience avec les méthodes d'apprentissage profond pour la classification d'images (fine-grained classification)
- Connaissances ou intérêt pour l'entomologie, la biodiversité et/ou la taxonomie, optionnels mais recommandés

Pour postuler, faites parvenir votre CV et une lettre de présentation à [andre-philippe.drapeaupicard@montreal.ca](mailto:andre-philippe.drapeaupicard@montreal.ca)



## **IA internship - Development of a visual recognition algorithm to identify butterfly species from photographs**

Duration: 8 months (with a possible 4 month extension)

Wage: 30 000 \$ (if Mitacs grant request is accepted)

Location: work from home, remote access to servers

Supervision: David Rolnick (McGill University, Mila) and Maxim Larrivée (Insectarium de Montréal, IRBV)

Posting period: 5 janvier 2022 - 21 janvier 2022

Context:

[eButterfly](#) is a participatory science program launched in 2012 which aims at documenting the biogeography of Lepidoptera. Initially limited to Canada and the United States, the program was extended in 2021 to Mexico, Central America and the Caribbean. To date, almost 500,000 observations have been submitted, representing 1,047 species. By 2022, the goal is to extend the program to the whole world.

Since 2020, the eButterfly team has been collaborating with AI researchers (Mila, Centrale Supélec) to develop a visual recognition algorithm to assist in the identification of submitted observations. A model was trained using images available on eButterfly and other platforms like GBIF, and incorporating metadata about where and when the images were taken. The model was called [eButterflyAI](#).

Internship description:

The overall objective of the project is to develop a tool to automatically identify global butterflies from images of living/pinned specimens submitted to eButterfly. This tool will facilitate and accelerate the identification and validation of new observations submitted by participants.

The first step will be to aggregate images from different sources to obtain a dataset in which more than 18,000 classes (species) are well-represented. Image sources will include eButterfly,



Institut de recherche  
en biologie végétale

GBIF, Leps by Fieldguide, BugGuide, and iNaturalist. A model will then be trained using these images and several “priors” such as the biology (phenology, geographical distribution) of the species. The model should be hierarchical and return a result at family-, genus- and species-level.

The photographed butterflies are expected to be identified at the family level with a high success rate (> 95%). In terms of genus and species, the success rate should be greater than or equal to 80% in the majority of cases. A lower rate would initially be acceptable for classes for which few images are available.

Required profile:

- PhD student
- Experience in deep learning methods for fine-grained image classification
- Knowledge or interest in entomology, biodiversity, and/or taxonomy is optional but preferred

To apply, please send your resume and a cover letter to

[andre-philippe.drapeaupicard@montreal.ca](mailto:andre-philippe.drapeaupicard@montreal.ca)