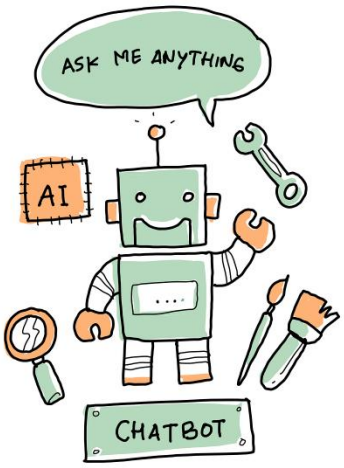

PEPR IA – Projet SAIF

Le département M2F est impliqué dans le projet SAIF (Sûreté de l'intelligence Artificielle Investiguée par les méthodes Formelles) du PEPR IA (https://project.inria.fr/saif/).
L'objectif est de développer des méthodes formelles pour rendre l'Intelligence Artificielle plus sûre. Ce projet sur 4 ans a débuté au 1er octobre 2023.

En savoir plus en consultant l'article ici.

L'IA (Intelligence Artificielle) avec plus particulièrement les récents progrès en ML (apprentissage automatique) fait désormais partie de notre paysage quotidien.
De nombreuses applications de notre société en sont équipées, notamment dans le domaine de la santé, l'énergie, les transports.

La révolution du ML a changé la façon dont les logiciels sont développés, offrant des performances au-delà de la programmation explicite traditionnelle.
Cependant des lacunes demeurent quant à sa fiabilité et son explicabilité. Le but de ces recherches est donc de la rendre plus sûre, d'être aussi plus transparent, afin qu'elle puisse être largement acceptée.
Assurer la fiabilité des logiciels n'est pas un objectif nouveau : plusieurs approches statistiques et des méthodes formelles (FM) ont été développées au fil des décennies pour traiter de ces problèmes. La communauté FM a passé ce temps à construire les fondations mathématiques et logiques pour des outils logiciels efficaces qui peuvent intervenir à toutes les étapes du cycle de vie des logiciels, de la spécification et du développement au déploiement et la maintenance. En fournissant des outils pour atteindre des normes de fiabilité plus élevées, la communauté FM a rendu le monde plus sûr, comme en témoigne le nombre croissant de normes qui exigent l'utilisation de FM dans le développement de logiciels critiques.
Cependant, les approches FM traditionnelles parviennent peu à répondre aux défis posés par les récents progrès de l'IA. La plupart des défis liés à l'utilisation de FM pour les systèmes basés sur l'apprentissage automatique (ML) découlent de leur nature stochastique et incertaine, de leur grande taille, de la difficulté de leur spécification, de leur nature monolithique, de leur manqué d'interprétabilité et de la profusion de différentes architectures qui créent un champ d'IA toujours plus large.
Pour relever ces défis, SAIF est organisé autour de trois objectifs, articulés le long de trois des principales étapes du cycle de vie traditionnel du développement de logiciels : la spécification, la conception et la validation des systèmes basés sur le ML. Chacun de ces objectifs comporte ses propres sous-objectifs, naturellement interconnectés entre eux. Ces sous-objectifs visent à pousser l'actuel état de l'art, par exemple à travers les artefacts d'IA traités (en particulier des architectures complexes telles que les réseaux de neurones graphiques et récurrents, peu explorés aujourd'hui), à travers les domaines d'application (en étendant aux types de données

non-images tels que les séries temporelles et la spectrométrie) et à travers les méthodes sous-jacentes aux outils d'amélioration de la fiabilité ciblés (comme les algèbres tropicales).
L'IA a déjà montré un grand potentiel pour notre société, mais pour réaliser pleinement son impact positif, il est essentiel de garantir rigoureusement sa sûreté. Cela nécessite un changement significatif dans la façon dont nous utilisons les FM pour traiter les problèmes modernes de l'IA.
L'objectif de SAIF est d'utiliser la vaste connaissance accumulée au fil des décennies dans les FM pour les repenser et aborder les nouvelles préoccupations en matière de fiabilité soulevées par le renouveau de l'IA. Grâce à la synergie d'un consortium diversifié avec une expertise complémentaire, nous visons à aider la société à se rapprocher d'un état où elle peut bénéficier des réalisations de l'IA sans en subir de fâcheuses conséquences.

Mots-clés : Program Synthesis, Programmatic Reinforcement Learning, Reactive Synthesis, Model Learning.

Contact : Nathanaël Fijalkow nathanael.fijalkow@labri.fr