Project de recherche doctoral

Directeur de thèse : Haralampos-G. Stratigopoulos

Laboratoire : LIP6

Domaine scientifique : Sciences et technologies de l'information et de la communication

Thématique CNRS : Intelligence Artificiel

Titre: Architectures matérielles fiable pour l'Intelligence Artificielle de confiance

Résumé: Aujourd'hui, les algorithmes d'Intelligence artificielle (IA), et plus particulièrement les réseaux neuronaux profonds (DNN), s'exécutent généralement dans le cloud sur des clusters de CPUs et GPUs. Pour pouvoir exécuter des algorithmes d'IA localement sur des systèmes embarqués, des implémentations matérielles efficaces pour l'IA (HW-AI) sont nécessaires. Cependant, comme tous les composants électroniques utilisés dans les architectures matérielles classiques, l'HW-AI est sujette à des pannes matérielles dues aux défauts de fabrication, au vieillissement des composants électroniques et aux perturbations environnementales. Bien que l'IA soit dotée d'un certain niveau de résilience aux pannes, celles-ci peuvent toujours affecter sérieusement l'inférence des DNN sur l'HW-AI et mener à des résultats complétement erronés. Par conséquence, des échecs de prédiction apparaissent, limitant sérieusement l'exécution de l'application. En outre, si le matériel est compromis, toute tentative d'explication des décisions de l'IA risque d'être peu concluante ou trompeuse. L'un des aspects négligés dans l'état de l'art est l'impact que les pannes matérielles peuvent avoir sur l'exécution de l'application et les décisions de l'HW-AI. Cet impact est d'une importance significative, en particulier lorsque l'HW-AI est déployée dans des applications critiques, telles que la robotique, l'aérospatiale, la santé et la conduite autonome. Ce projet inclura l'impact de la fiabilité de l'HW-AI sur la fiabilité, la confiance et l'explicabilité des décisions de l'IA. Les approches typiques, telles que les tests en ligne et la redondance matérielle, ou même le réentraînement, sont moins appropriées pour l'HW-AI en raison des coûts très élevés en termes de surface et consommation énergétique. En effet, les DNN ont des architectures matérielles complexes, avec en plus des bases des données d'apprentissage très grandes. Le projet abordera ces limitations en exploitant les particularités des architectures HW-AI pour développer des stratégies de fiabilité efficaces et en même temps peu coûteuses. Pour y parvenir, le projet développera ainsi des modèles de défauts et effectuera une analyse de défaillance des HW-AI pour étudier leur vulnérabilité dans le but « d'expliquer » le HW-AI. Expliquer le HW-AI signifie s'assurer que le matériel est sans erreur et qu'il ne compromet et ne biaise pas la prise de décision de l'IA. À cet égard, le projet vise à apporter de la confiance dans la prise de décision basée sur l'IA en expliquant le matériel sur lequel les algorithmes d'IA sont exécutés.