# Deep Machine Learning for Cyber Defence

## (STO-TR-IST-163)

# Executive Summary

Cyber threats grow increasingly pervasive. Recent high-profile intrusions illustrate how surreptitious cyberspace effects can challenge the 21st century's strategic international security landscape. The growing reliance upon digital technology in every economic sector and aspect of human life strongly suggest this trend will continue. NATO Allies are responding with increasingly robust security and defence of the cyber landscape, especially as it intersects with military systems, platforms, and missions.

The demand for increased resilience and robustness has accelerated the exploration and adoption of Artificial Intelligence technology, i.e., techniques that enable computers to mimic human intelligence, for cyber defence. Deep Machine Learning (DML) is one such state-of-the-art technique which demonstrates considerable potential in cybersecurity as well as many other application domains. Deep Machine Learning can enhance cyber resilience with defences that evolve with threats over time and reduce the overall burden of manual data analysis by human experts. Deep Machine Learning facilitates faster responses, most especially with ample and sufficient training. Some possible consideration includes adversarial examples within training and model development in building or generating data models.

This technical report takes initial steps in consolidating NATO-wide knowledge in the field of cyber defence applications of DML. It further identifies gaps between current solutions and military needs and structures the pursuit of promising cyber defence applications of DML for the military domain accordingly. The research group, with the embodiment of the technical report at a core, examines the National Institute of Standards and Technology security guidelines from the perspective of Malware Detection, Event Management, Information Management, Vulnerability Management, Software Assurance, Asset Management, Licence Management, Network Management, and Configuration Management.

The report examines the intricate utility of DML, practical implementations as well as open challenges. The Research Task Group comprises experts across the fields of data science, machine learning, cyber defence, modelling & simulation, and systems engineering. Researchers and practitioners consider aggregation of data, characterization of data, the need to share data, and the sharing of data models, or the generators thereof. These factors, including how data will be processed, trained, accessed, and related techniques such as transfer, or federated learning are also considered.

# Apprentissage automatique profond pour la cyberdéfense

## (STO-TR-IST-163)

# Synthèse

Les cybermenaces sont de plus en plus omniprésentes. De récentes intrusions très médiatisées illustrent de quelle façon le cyberespace peut subrepticement bouleverser le paysage international stratégique de la sûreté. La confiance croissante accordée à la technologie numérique dans tous les secteurs économiques et aspects de la vie humaine suggère fortement que cette tendance perdurera. Les Alliés de l'OTAN y répondent par une sûreté et une défense robustes du paysage cybernétique, en particulier lorsque celui-ci recoupe les systèmes, plateformes et missions militaires.

La demande de renforcement de la résilience et de la robustesse a accéléré l'exploration et l'adoption des technologies d'intelligence artificielle, autrement dit, des techniques qui permettent aux ordinateurs d'imiter l'intelligence humaine, pour la cyberdéfense. L'apprentissage automatique profond (DML) est l'une de ces techniques de pointe qui montre un potentiel considérable en cybersécurité, comme dans beaucoup d'autres domaines d'application. L'apprentissage automatique profond peut favoriser la cyber-résilience en faisant évoluer la défense avec les menaces et en réduisant la charge générale d'analyse manuelle des données par des spécialistes humains. L'apprentissage automatique profond accélère la réponse, plus particulièrement avec un entraînement suffisant et de grande ampleur. Il peut être envisagé de fournir des exemples d'adversaire pendant l'entraînement et le développement des modèles, pour établir ou produire des modèles de données.

Le présent rapport technique constitue un premier pas vers la consolidation des connaissances de l'OTAN en matière d'applications du DML à la cyberdéfense. Ce rapport identifie l'écart existant entre les solutions actuelles et les besoins militaires et structure en conséquence la recherche d'applications prometteuses du DML à la cyberdéfense dans le domaine militaire. Le groupe de recherche, dans l'optique du rapport technique, a examiné les lignes directrices du National Institute of Standards and Technology en matière de sûreté, sur le plan de la détection des logiciels malveillants, la sûreté des logiciels et la gestion des événements, des informations, des vulnérabilités, des actifs, des licences, du réseau et de la configuration.

Le rapport étudie l'utilité complexe du DML, les applications pratiques de celui-ci et les défis lancés. Le groupe de recherche se compose d'experts en science des données, apprentissage automatique, cyberdéfense, modélisation et simulation et ingénierie des systèmes. Les chercheurs et praticiens étudient l'agrégation et la caractérisation des données, le besoin de partager les données et le partage des modèles de données ou de leurs générateurs. Ces facteurs, notamment le mode de traitement des données, l'entraînement du système, l'accès aux données et les techniques liées telles que l'apprentissage par transfert ou l'apprentissage fédéré, sont également étudiés.