« Analyse prédictive, *Big data*

et ciblage large spectre »

LCL ANGELOTTI Benoît

25ème Promotion de l'Ecole de Guerre

Directeur de mémoire : ICETA E.PFANNSTIEL

Remerciements

Je remercie le directeur de ce mémoire, l'ingénieur Pfannstiel pour ses orientations, le colonel Martin et le colonel Kirsch, pour leurs conseils avisés.

J'ai une pensée toute particulière pour toute ma famille et notamment pour mes enfants afin de saluer leur patience et compréhension de tous les jours.

SOMMAIRE

Table des matières

Résumé

Les nouvelles technologies liées à l'exploitation des données de masse appelées *big data* ainsi que les algorithmes d'analyse appelé prédictifs ont pris une importance croissante dans de nombreuses activités des secteurs civils et privés.

Le traitement de flux de données massifs ainsi que leurs exploitations à grande vitesse a permis des progrès significatifs dans les domaines de la logistique ou du renseignement. En parallèle, les algorithmes prédictifs ont optimisés le domaine de la sécurité intérieure et ouvert des perspectives dans le monde de la finance. Ces deux disciplines se sont rencontrées permettant d'atteindre un point d'inflexion après lequel toutes les disciplines se trouvent optimisées via l'exploitation de flux de données important, l'utilisation d'algorithmes prédictifs associés en systèmes apprenants.

Du côté des armées et de la Défense, des besoins sont apparus devant la complexité des crises à traiter, le tempo et l'évolution des conflits ainsi que les contraintes de ressources.

Le domaine du ciblage large spectre en particulier, nécessite une connaissance fine, à jour et immédiate des interactions entre les différents facteurs qui conditionnent le système ennemi ou adverse. Dans ce domaine, l'analyse prédictive pourrait apporter une forte plus-value. Cependant, appliquer les nouvelles technologies au ciblage large spectre de niveau stratégique comporte de nombreux obstacles qui révèlent de nombreuses contraintes. Il s'agit donc d'identifier les points qui bénéficieraient rapidement de l'utilisation des nouvelles technologies de la donnée. En cela, les cycles du ciblage semblent être des leviers d'applications les plus rentables. Le domaine des opérations spéciales constituer un laboratoire adapté.

La portée des nouvelles technologies révèle la problématique de l'intelligence artificielle. La puissance des outils ne doit pas retirer la responsabilité de la décision humaine. L'affrontement militaire n'est pas à comparer avec la recherche de profit ou de parts de marché. L'investissement n'est pas neutre, il importe de se doter de compétences humaines adaptées mais également de s'appuyer sur des laboratoires d'innovations. Les structures de commandement nécessitent enfin de s'adapter.

Les technologies du numériques ont donc un réel intérêt dans le domaine du ciblage large spectre. Cependant, l'investissement dans ces disciplines s'avère couteux et doit être effectué par briques successives. L'intérêt doit être double, obtenir des résultats à courts termes et permettre la construction d'un système global pertinent de niveau stratégique.

Abstract

New technologies related to the exploitation of mass data called big data as well as predictive analysis algorithms have become increasingly important in many civil and private sector activities.

The processing of massive data flows and their high-speed operations has led to significant progress in the areas of logistics or intelligence. In parallel, predictive algorithms have optimized the field of internal security and opened up prospects in the world of finance. These two disciplines met to reach a point of inflection after which all disciplines are optimized through the exploitation of important data flows, the use of predictive algorithms associated in learning systems.

On the side of the armies and the Defense, needs appeared in front of the complexity of the crises to be treated, the tempo and the evolution of the conflicts as well as the constraints of resources.

The field of broad spectrum targeting in particular, requires a fine, up-to-date and immediate knowledge of the interactions between the various factors that condition the enemy or adversary system. In this area, predictive analysis could bring a strong added value. However, applying new technologies to broad-based strategic-level targeting has many hurdles that reveal many constraints. It is therefore necessary to identify the points that would quickly benefit from the use of new data technologies. In this, targeting cycles seem to be the most profitable levers of applications. The field of special operations is a suitable laboratory.

The scope of new technologies reveals the problem of artificial intelligence. The power of tools must not take away the responsibility of human decision. Military confrontation is not to be compared with the search for profit or market shares. The investment is not neutral, it is important to have adapted human skills but also to rely on innovation laboratories. Command structures finally need to adapt.

Digital technologies are therefore of real interest in the field of broad-spectrum targeting. However, investing in these disciplines is expensive and must be done in successive blocks. The interest must be twofold, obtain short-term results and allow the construction of a relevant global system of strategic level.

I - Introduction

> ***« Les principes de la guerre édictés par Foch – « l'économie des moyens », « la concentration des efforts » et « la liberté d'action » – restent une matrice pour penser la manière de conduire la guerre. Mais à ces principes, je crois qu'il faut ajouter la surprise. Non celle que l'on subit, mais celle que l'on impose. »***
>
> ***Général d'armée Pierre de Villiers.***

Les conflits modernes ont mis en évidences des modes d'actions ennemis fortement inspirés des techniques de guérillas faisant la part belle à la dissimulation, à la surprise. L'ennemi –et c'est particulièrement vrai lorsqu'il emploie des techniques terroristes– cherche à se fondre dans la population, dans l'environnement. Il n'hésite pas à utiliser des procédés perfides pour cacher ses intentions. Les grandes puissances sont enclines à faire appel à des procédés dits de guerre hybride dont les pratiques font une fois de plus appel à la dissimulation et à la manipulation. Tant et si bien que les guerres d'aujourd'hui sont devenues avant tout des guerres d'information et de renseignement.

En parallèle, le numérique et les technologies numériques ont commencé à inonder le monde qu'il soit industriel, média, du quotidien, du grand public ou des recherches de pointe. Le monde militaire n'échappe pas à ce phénomène et à cette nouvelle révolution technologique, la première révolution aussi formidable depuis la révolution industrielle du XIX^{ème} siècle. La puissance des supercalculateurs autorise aujourd'hui d'atteindre des niveaux d'équation qui permettent de prédire la météo plusieurs jours à l'avance avec des taux de probabilité supérieurs à 80%.

Enfin, les disciplines qui développent les algorithmes progressent de façon continue. Il devient alors tentant de considérer ces nouvelles possibilités comme des outils en mesure de répondre à certaines contraintes et difficultés des conflits modernes. Au regard de la complexité, du coût potentiel et des risques induits par le développement des technologies associées au numérique et au calcul, une étude semble nécessaire pour présenter les avantages potentiels, les contraintes induites, dégager des axes d'efforts essentiels et fixer des priorités dans l'acquisition et la mise en œuvre de système et enfin de proposer un ou plusieurs modèles d'application de ces nouvelles techniques à la prévention et à la résolution des conflits.

L'enjeu de cette étude consiste à identifier si les techniques dites prédictives ou relatives au *big data* constituent une rupture dans la stratégie ou à défaut un *game changer* dans l'approche de la prévention ou de la résolution des conflits. Un autre enjeu sous-jacent demeure l'investissement qui pourrait être consenti au profit du développement de l'intelligence artificielle.

L'analyse prédictive est aujourd'hui utilisée en majeure partie au profit des techniques de vente ou plus largement de la discipline appelée marketing. On comprendra sous ces termes la définition des stratégies générales des entreprises pour augmenter et optimiser leurs ventes en dégageant un maximum de bénéfice. Le domaine du *big data* est essentiellement employé dans le domaine industriel ou des services pour optimiser les procédés de production ou de distribution. Il permet également d'identifier des clients potentiels.

La lutte contre la criminalité, le crime organisé et le terrorisme utilisent également des logiciels d'analyse prédictive ainsi que des moteurs de recherche sur de grandes bases de données. Les études portées sur ces premières applications confirment leurs pertinences. Elles annoncent également un profond essor du domaine et envisagent des possibilités d'optimisation sans précédent consécutives des progrès constants des technologies numériques.

Cependant ces outils n'ont jamais été utilisés ou étudiés dans la perspective d'un travail de ciblage de niveau stratégique. Aujourd'hui, le Ciblage Large Spectre (CLS) fait partie intégrante de la réflexion de ce niveau. L'intérêt de ce document est d'évaluer la pertinence de l'emploi de ces technologies au profit du CLS. Les études prospectives font plus état de questions éthiques voire philosophiques portant sur le rôle et l'importance de l'intelligence artificielle dans les sociétés humaines. Ces productions qui relèvent plus de l'anticipation voire de l'activité d'essayistes que de la prospective technologique n'abordent pas la portée ni les difficultés liées à l'utilisation de puissants calculateurs associés à des algorithmes complexes et sur des masses de données gigantesques au profit de la décision stratégique. Le sujet utilise mais dépasse la simple notion de simulation. En effet, l'association des outils devraient idéalement conduire à la proposition de concepts d'opération et permettre d'évaluer les conséquences de la variation d'un paramètre.

Le travail de réflexion mené dans ce document s'appuie essentiellement sur des articles électroniques en ligne sur internet qui traitent de l'analyse prédictive, des algorithmes et du *big data*. De nombreux exemples de systèmes existants sont présentés dans leur utilisation ou dans leur mise en œuvre.

Ces exemples constituent une source d'évaluation du domaine, même si la majeure partie de ces systèmes sont développés dans un cadre industriel ou financier.

Il repose également sur des ouvrages qui considèrent le développement de l'intelligence artificielle et des systèmes apprenants. L'essentiel de ces ouvrages est issu du monde anglo-saxon. Plusieurs mémoires et travaux de réflexions portent sur l'utilisation de systèmes prédictifs dans le cadre de l'optimisation de la sécurité publique et de la lutte contre la criminalité. La totalité de ces travaux a été rédigée par des anglo-saxons. Enfin, un nombre significatif d'entretiens libre avec des opérateurs de systèmes exploitant des données, des architectes de systèmes d'information, des concepteurs de projets informatiques portant sur des stratégies de ciblage marketing. La production doctrinale française et alliée constitue le socle des réflexions liées au ciblage. Une partie significative des méthodes appliquées par les agences de renseignement a constitué un modèle d'inspiration même si les techniques aperçues ne seront pas dévoilées dans un document non classifié.

La réflexion a été confrontée à trois difficultés majeures. Tout d'abord, la faiblesse de la production d'ouvrage français sur ces domaines assez récents. L'essentiel de la production hexagonale se concentre sur le domaine cyber qui regroupe la cyberguerre, le cyber-terrorisme et la cybercriminalité. Il n'a pas été possible de traiter ces aspects qui certes affectent le sujet étudié mais qui constituent un domaine plus vaste et très distinct du cœur de la réflexion menée. Il s'éloigne ainsi de la problématique de l'emploi et des perspectives de l'outil numérique dans le domaine du ciblage large spectre, pour étudier un champ de guerre différent. En effet, *a priori* le domaine des réseaux ne concerne pas cette étude. Nous nous efforcerons de le traiter sans excès. Ensuite, le sujet est le plus souvent abordé autour de son aspect technique. L'essentiel de la littérature y faisant référence fait la part belle aux algorithmes, aux équations et aux modèles mathématiques. L'objectif de ce mémoire n'est pas de présenter des solutions techniques ni d'aborder l'aspect scientifique fondamental de la technologie. Les conclusions et les problématiques générales sont évoquées en revanche dans une perspective plus large qui inclue les qualifications nécessaires à la mise en œuvre et le coût des systèmes étudiés. Enfin, le manque d'homogénéité des domaines en question lié aux faits que ces disciplines soient assez neuves et que le développement numérique soit fulgurant complique la compréhension des problématiques.

En effet, les définitions et les termes usuels divergent et n'ont pas la même portée en fonction des pays ou des domaines dans lesquels les technologies sont appliquées.

Tout en tenant compte de ces difficultés, l'exposé de ce document cherche à déterminer si la nature et la portée du ciblage large spectre peut être fondamentalement modifiée par la mise en œuvre de systèmes utilisant l'analyse prédictive et les technologies relatives au *big data*. L'essor extraordinaire de ces domaines liés au numérique constitue-t-il potentiellement une rupture dans la planification et la conduite stratégique ? Ce document cherche en particulier à évaluer la rentabilité, les contraintes et les risques induits par la mise en œuvre des systèmes numériques. Il recherche également la portée potentielle que peut prendre la mise en œuvre de ces systèmes et les opportunités nouvelles ou les éventuels risques dans la conduite des conflits et son implication dans la chaine de décision stratégique.

Une première partie se concentrera sur l'analyse des systèmes existants. Puis, une deuxième partie observera l'utilisation qui peut en être faite dans le domaine du ciblage en soulignant les problématiques majeures. Enfin, une dernière partie dégagera des perspectives et présentera des recommandations en particulier en termes d'organisation, de compétences nouvelles et d'investissement.

II – Analyse des systèmes existants

A. Définitions

Big data : « Les *big data* ou mégadonnées désignent l'ensemble des données numériques produites par l'utilisation des nouvelles technologies à des fins personnelles ou professionnelles. Cela recoupe les données d'entreprise (courriels, documents, bases de données, historiques de processeurs métiers...) aussi bien que des données issues de capteurs, des contenus publiés sur le web (images, vidéos, sons, textes), des transactions de commerce électronique, des échanges sur les réseaux sociaux, des données transmises par les objets connectés (étiquettes électroniques, compteurs intelligents, smartphones...), des données géolocalisées, etc. »[1]

Analyse prédictive : « Le terme analyse prédictive rassemble de nombreuses technologies d'analyse de données et autres techniques statistiques. La principale technique est l'analyse de régression, permettant de prédire les valeurs reliées de multiples variables en se basant sur la confirmation ou l'infirmation d'une affirmation

[1]FUTURA TECH. (L'expression « *Big Data* » date de 1997 selon l'*Association for Computing Machinery*). https://www.futura-sciences.com/tech/big-data/ de 2017

particulière. Les analyses prédictives visent à reconnaître des patterns dans les données pour la probabilité d'un projet. »[2]

Data mining : « le terme *Data Mining* désigne l'analyse de données depuis différentes perspectives et le fait de transformer ces données en informations utiles, en établissant des relations entre les données ou en repérant des patterns. Techniquement, le *Data Mining* est le procédé permettant de trouver des corrélations ou des patterns entre de nombreuses bases de données relationnelles.

Le Data Mining repose sur des algorithmes complexes et sophistiqués permettant de segmenter les données et d'évaluer les probabilités futures. Le *Data Mining* est également surnommé *Knowledge Discovery in Data*. »[3]

Machine learning : « L'apprentissage statistique (*Machine Learning*) est un type d'intelligence artificielle qui confère aux ordinateurs la capacité d'apprendre sans être explicitement programmés. Cette technologie s'appuie sur le développement de programmes informatiques capables d'acquérir de nouvelles connaissances afin de s'améliorer et d'évoluer d'eux-mêmes dès qu'ils sont exposés à de nouvelles données.

Le processus d'apprentissage automatique s'apparente à celui de l'exploration de données (*data mining*). En effet, il s'agit, dans les deux cas, d'analyser les données à la recherche de schémas récurrents. Cependant, au lieu d'extraire les données pour les soumettre à un traitement humain (comme c'est le cas dans les applications de *data mining*), l'apprentissage automatique utilise ces données pour améliorer la compréhension du programme lui-même. »[4]

Deep learning :

Data analyst : Un analyste de données (*data analyst*) a pour mission d'agréger des données venant de multiples sources, de les analyser et d'en extraire des informations permettant à son entité de mieux piloter ses activités et d'anticiper ses futurs besoins. Une partie essentielle de son travail est de restituer les conclusions de ses analyses aux autres services de son entité.[5]

[2] LE BIG DATA, selon Allison Snow, Senior Analyst du B2B Marketing chez Forrester de 2017
[3] http://www.lebigdata.fr 2017
[4] http://www.lemagit.fr de 2016
[5] http://www.blogdumoderateur.com/metiers-data-regionsjob

Il est donc capable d'interroger des sources de données pour en faire des rapports et des visualisations graphiques (graphes, camemberts, histogrammes, etc.). Il a une compréhension forte du domaine métier dans lequel il opère.

Ce qui lui permet de mieux communiquer avec les gens du métier. Pour mieux explorer les données, un *data analyst* est généralement à l'aise avec les outils statistiques. Toutefois, il n'est pas forcément aussi competent techniquement qu'un ingénieur logiciel pour traiter les grands volumes de données (*big data*)[6]

Data scientist : Le data scientist a la même mission que le data analyst, exploiter et valoriser les données, mais dispose de compétences différentes. En un sens, le data scientist se substitue au data analyst quand l'analyse des données devient plus complexe et exige la maîtrise de techniques et outils plus pointus. Cela peut être le cas quand le volume des données devient très grand (*big data*), quand les données doivent être traitées très rapidement (pour des applications en temps réel, par exemple), quand la nature même des données exige des traitements spécifiques (traitement du langage ou des images), ou quand le problème à résoudre nécessite une modélisation poussée relevant davantage de la R&D que de la restitution de statistiques.[7]

Data engineer : Un data engineer est quelqu'un ayant un background technique en développement logiciel. Il peut être un software engineer qui s'est reconverti dans le big data. Les data engineers vont mettre en place des systems de big data pour traiter ces dernières. Ils opteront pour des outils de stockage performants et se baseront sur des moteurs de traitement de données rapide (comme Hadoop ou spark) pour traiter convenablement ces grands volumes de données. Mes data engineer vont collecter, transformer les données des différentes sources. Ce travail permettra d'avoir des données prêtes pour qu'on leur applique des techniques de machine learning. Le travail d'un data engineer est donc de préparer le terrain pour qu'un data scientist puisse se server des données propres pour en tirer des tendances. Il s'occupe du côté applicatif permettant le travail des data scientist. Il développe et entretient les systèmes de collecte, stockage et mise à disposition des données

B.I. developer : Les *Business Intelligence developer* vont mettre en place des outils d'informatique décisionnelle pour les besoins de l'entreprise. Ces outils se présentent généralement sous forme de *data warehouses, datamart*, ainsi que de bases de données

[6] mrmint.fr de 2017
[7] https://www.blogdumoderateur.com/metiers-data-regionsjob du 12/12/2017

multidimensionnelles construits à partir d'agrégation de données en provenance de plusieurs bases. Ces bases de données multidimensionnelles et data *warehouses* sont par la suite utilisées par les *B.I. developer* pour construire des tableaux de bord et des rapports utiles pour les manageurs et les décideurs.

Les B.I. developer ont généralement une connaissance métier moindre que les *data analyst.* Cependant, ils maitrisent mieux techniquement l'interfaçage avec les différentes sources de données.

B. Analyse prédictive

a. Dans la finance

L'irrationalité des marchés financiers a toujours été soulignée par les spécialistes. Au sein de ces marchés, les agences financières doivent identifier les secteurs dans lesquels investir et ceux sur lesquels il importe de miser sur la dépréciation.

A l'échelle planétaire, de grands profits peuvent être tirés des différences de prix d'un produit selon qu'il se trouve en un lieu ou un autre de la planète ou du marché. La vitesse d'exploitation de l'information est aujourd'hui de l'ordre de la nanoseconde. Les masses financières dégagées par les actions financières, ont poussé la finance à investir lourdement dans l'analyse prédictive.

Le recueil et l'exploitation des données ainsi que l'utilisation d'algorithmes mathématique sont utilisés depuis plusieurs années dans la définition des stratégies financières. Cependant, l'évolution des capacités techniques et l'augmentation des flux de données ont permis d'atteindre un point d'inflexion qui fait que la plupart des sociétés bancaires ont maintenant recours à l'analyse prédictive pour augmenter leur performance.

Les sociétés les plus performantes développent même des architectures de détection basées sur la pondération de certain facteurs qui permettent de lancer des alertes ou d'identifier des opérations à forte rentabilité.

De manière générale, l'utilisation des données dans la finance renseigne les preneurs de décision et pilote la planification financière pour optimiser les opérations, prédire les cotations et remporter de nouveaux marchés.

b. Dans la sécurité intérieure

L'utilisation de modèles prédictifs dans la sécurité intérieure a montré de premiers résultats. Les logiciels, pour leur majeure partie ont été mis en œuvre dans de grandes villes américaines (Santa Cruz, Shreveport). La France n'est pas absente de cette innovation et a déjà mené certaines expérimentations en parallèle du logiciel ANACRIM. Il parait nécessaire de préciser que ce logiciel n'est pas un système prédictif à proprement parler mais plutôt un outil qui permet de mettre en valeur des liens complexes et d'identifier certaines associations. Il n'en demeure pas moins pertinent en termes d'algorithmie.

La littérature révèle plusieurs écoles de pensée au sujet de la capacité des formules mathématiques à prédire où et quand un crime particulier aura lieu. L'opinion est partagée entre ceux qui professent que les algorithmes sont réellement capables de faire des prédictions et ceux qui stipulent que les algorithmes se contentent de fournir des zones de probabilité croissante où certains crimes peuvent avoir lieu. Il est communément admis que les crimes et les troubles à l'ordre public se concentrent dans des zones restreintes ou points chauds.

Des études ont prouvé que renforcer l'activité des forces de l'ordre sur ces zones peut être efficace et prévenir ou réduire l'activité criminelle. Les actions proactives du type patrouilles dirigées ou renfort de la police routière sont associées à des réductions significatives de la délinquance sans déplacement notable du problème ailleurs. La police ciblée est fondée sur la supposition que le meilleur indicateur de performance demeure l'indicateur historique. La police ciblée est donc dépendante des données historiques dans le processus décisionnel et l'allocation des moyens de sécurité intérieure.

De nombreux programmes d'analyse prédictive basés sur la géolocalisation reposent sur des théories de répétitions de cas de victimes et sur des procédés mathématiques qui permettent de créer des modèles de répétitions afin de produire des cartes de zones où les crimes et la délinquance ont le plus de chance de se produire. Ces modèles dépendent d'un procédé à nuages de points pour analyser les données et prédire les crimes ou délits à venir.

c. Dans l'analyse électorale

Il est généralement admis que l'ensemble des outils du marketing commercial sont utilisés dans les campagnes électorales. Il s'agit généralement de spots télévisés, d'affiches, de mailings, de marketing téléphonique. Depuis 2008 et la campagne de Barack Obama à l'élection présidentielle américaine, l'approche a considérablement évolué avec la prise en compte de la maîtrise des bases de données. Jusque-là, les campagnes s'appuyaient sur des données chiffrées précises, parfois complexes qui dessinaient le paysage politique. Une connaissance fine de l'opinion et des enjeux était ainsi permise. Les campagnes électorales comme les études politiques étaient largement fondées sur cette analyse. La nouveauté réside dans la collecte d'un nombre considérable de bases de données aussi bien de nature politique que commerciale et surtout dans leurs connexions et leurs rapprochements. Cette interconnexion de multiples données disparates permet de disposer d'informations inédites et jusque-là inaccessibles sur l'ensemble des citoyens et leurs comportements. Le fonctionnement de la campagne d'Obama a ainsi été en majeure partie dominé par des décisions issues de l'exploitation de ces mégadonnées.

Le logiciel PREVIOO constitue un exemple intéressant : « PREVIOO est un logiciel en ligne de gestion et d'animation de campagnes électorales. Il s'adresse aux candidats à une élection et à leurs équipes de militants mais pas seulement. Des élus via leur association citoyenne s'en servent aussi pour rester en contact avec leurs administrés en collectant des données, en les hiérarchisant, en les analysant pour mieux piloter leurs actions »[8]. Il possède comme fonctionnalités principales de déterminer et géo-localiser ses électeurs bureau de vote par bureau de vote ; d'organiser et hiérarchiser son équipe et ses militants, d'assigner des missions(porte à porte, parrainage, collage d'affiche,…) tournées vers les électeurs ciblés ; d'avoir un retour terrain synthétique de l'avancement de la campagne ; de collecter des données électeur par électeur, les hiérarchiser, les analyser automatiquement ; de rester en contact avec ses électeurs en envoyant des Mails et SMS ciblés.

[8] HTTP://www.previoo.com, octobre 2016

Cette analyse croisée des données permet d'identifier, par bureau de vote, quartier par quartier, les électeurs susceptibles d'évoluer et sur lesquels des actions ciblées méritent d'être conduites. Ainsi en déterminant les électeurs indécis ou modérés, un nouveau marketing politique a vu le jour. Le micro-ciblage est devenu possible.

En utilisant le porte à porte, on ne touche ainsi qu'une fraction du corps électoral mais qui peut s'avérer déterminante en fonction des stratégies adoptées. Ainsi, dans le cas de l'élection américaine, ce sont les états indécis, les « swing states » qui constituent l'enjeu principal des élections présidentielles. Les zones cruciales peuvent être identifiées via le croisement des données notamment démographiques, commerciales, politiques, etc. La différence repose sur deux points. Tout d'abord, lorsqu'une personne est abordée en porte à porte, une bonne partie de ses habitudes et de son comportement est connu (consommation, voyages récents, relations familiales, réseaux internet, etc.). Ensuite, l'utilisation des profils et des algorithmes prédictifs permettent d'affiner le message qui lui sera délivrer en fonction de simulations de réactions effectuées au préalable.

En résumé, l'exploitation des données et leurs croisements permet de mieux connaitre l'électorat, de le cibler, de mieux connaitre les cibles, de prédire leurs réactions et donc le message à leur délivrer.

Ce modèle est peu à peu transposé en France. Même si le vivier en termes de bases de données est plus faible qu'aux Etats-Unis, d'importantes données commerciales et démographiques permettent de cibler de façon pertinente les électeurs.

C. Big data

a. Dans les processus logistiques et d'approvisionnement

Le processus d'exploitation des données à l'ère du *big data* suit plusieurs étapes successives qui s'avèrent interdépendantes. Il s'agit des étapes suivantes : filtrage des données, valorisation, exploitation, capitalisation et partage. Dans les entreprises, à l'issue de l'étape partage, les étapes de décision puis d'évaluation ou d'estimation des résultats concluent habituellement ce processus. Le filtrage permet d'effacer le bruit de fond des données parasites et d'éviter de forcer le trait de l'infobésité. Cela permet également d'éviter les doublons, de ne pas aggraver la portée des données peu fiables et de limiter la dissémination de données classifiées aux mauvais niveaux. La valorisation des données consiste à la stocker de façon utile et pratique à

l'exploitation, à ce niveau, la donnée est enrichie. L'exploitation consiste à analyser les données et donner du sens afin de générer de l'information à partir de données brutes enrichies.

La capitalisation consiste à noter et hiérarchiser les informations nées de la phase précédente. Le partage consiste à transférer l'information ou la rendre accessible par le ou les bonnes personnes ou bien les bons systèmes.

Dans la fonction de gestion de la chaîne logistique, on parle de degrés d'autonomie dans le processus de planification. Ce degré s'évalue entre la planification classique à partir des données jusqu'à l'automatisation totale qui ne laisse aucune option et ne permet pas de vérifier les recommandations des algorithmes. « La mise en place d'une chaîne logistique basée sur des algorithmes exige une certaine maturité de l'organisation, laquelle doit être préparée à intégrer et à s'appuyer régulièrement sur des systèmes. » rappelle le Dr. Ravi Prakash Mathur, directeur de la gestion de la chaîne logistique et responsable de la logistique et de la planification centrale chez *Dr. Reddy's Laboratories Ltd*. Mais le concept de chaîne logistique intelligente va encore plus loin, en incorporant des capacités d'auto-apprentissage de la machine pour optimiser la prise de décision propre à la chaîne logistique.

La compétitivité des entreprises dépend de l'efficacité de leur chaîne logistique. À l'avenir, leur compétitivité sera déterminée par la puissance de l'intelligence intégrée à leurs systèmes. Les plus compétitives seront celles dont la chaîne logistique apprendra avec la plus grande rapidité et la plus haute précision.

Un algorithme d'apprentissage automatique ou système apprenant est un jeu de données d'enseignement. L'ordinateur répond à chaque question en ajoutant chaque bonne ou mauvaise réponse possible au jeu de données. Ainsi, l'algorithme s'améliore et s'affine donc au fil du temps.

Dans les entreprises, les personnes en charge de la chaîne logistique se servent de l'intelligence intégrée pour examiner et modifier les prévisions, les plans de production ou les plans d'approvisionnement générés automatiquement.

L'introduction d'une boucle d'auto-apprentissage dans le système permet par exemple, à la machine d'analyser pourquoi sa recommandation a été remplacée manuellement et de vérifier ce paramètre au cycle suivant. Cette fonctionnalité est utile pour des

opérations de gestion, comme pour corriger des paramètres incorrects, modifier des normes ou répondre à la dynamique d'un marché en constante évolution.

Dans un premier temps, les systèmes doivent apprendre l'ordre des priorités suivi par les responsables à partir des scénarios de gestion émergents, et pas seulement à partir d'algorithmes d'optimisation.[9]

b. Dans le domaine du renseignement

Le domaine du renseignement est le premier concerné par l'info-obésité. Collecter des flux toujours plus importants de données grâce à une multitudes de capteurs dont la qualité ne cesse de s'améliorer n'est pas très utile si l'on ne dispose des moyens nécessaires pour trouver, vérifier et recouper la bonne information au moment opportun. Selon le général Jean-François Ferlet, Directeur du renseignement militaire (DRM), Cela peut même être « paradoxalement devenir un handicap ».

Une mission de drone Moyenne Altitude Longue Endurance (MALE) fournit environ 30 Térabits de données. Le flux d'images satellitaires transmis aux analystes de la DRM est déjà considérable. Or, dans un avenir proche, son volume sera multiplié par 10, voire 20. Ce principe concerne aussi le Centre de formation et d'emploi relatif aux émissions électromagnétiques (CF3E), du Centre interarmées de recherche et de recueil du renseignement humain (CI3RH) ou encore du Centre de recherche et d'analyse du cyberespace (CRAC). En effet, les programmes satellitaires MUSIS et CERES augmenteront significativement les capacités d'écoute, de surveillance et d'observation.

Cette tendance est la même aux États-Unis. « Si nous devions exploiter manuellement toutes les images satellites que nous nous attendons à recevoir au cours des 20 prochaines années, il faudrait embaucher huit millions d'analystes spécialisés dans l'imagerie », avait indiqué, en juin 2017, Robert Cardillo, le directeur de la National Geospatial-Intelligence Agency.

En mars 2017, « l'Intelligence Campus » a été créé sur la base de Creil. Ce campus a été décrit comme le « premier écosystème européen civil et militaire en traitement de la donnée » par Jean-Yves Le Drian, alors ministre de la Défense. L'idée est de réunir la communauté du renseignement (et en particulier la DRM) et des entreprises

[9] https://news.sap.com/france/2017/09/21/cas-dutilisation-de-lintelligence-artificielle-au-profit-de-la-chaine-logistique/ du 12 mars 2018

innovantes ainsi que des laboratoires spécialisés dans l'intelligence artificielle,le big data et le machine learning.

L'enjeu, pour la DRM, est donc de trouver des moyens pour analyser automatiquement ces flux de données. L'évolution des technologies fait partie des principaux problèmes structurants à anticiper.

Le général Ferlet est particulièrement attentif à ces questions : « Il faut qu'on se fasse aider par des outils d'intelligence artificielle qui vont nous aider à exploiter dans ce nuage de données l'information pertinente quand on en a besoin », même si « *cela ne remplacera pas les analystes* ». Il ajoute que « *Ce n'est pas tant une question de financement* » mais plutôt de « *solutions techniques* ».

Aux Etats-Unis, la communauté du renseignement mise aussi sur l'intelligence artificielle, avec l'appui de la Silicon Valley. La CIA, notamment, a au moins 137 projets relatifs aux traitements des données. Comme l'a expliqué Dawn Meyerriecks, son responsable du développement technologique, ces outils lui pourraient lui permettre de « prédire des évènements importants, politiques ou autres, en trouvant des corrélations entre des changements dans les flux de données et d'autres informations. »[10]

III – Des systèmes immédiatement applicables dans le domaine du ciblage

A. Caractéristiques du domaine du ciblage large spectre

Le ciblage est un processus décisionnel de sélection, de recherche, d'acquisition et de traitement des objectifs. Le ciblage est un processus comportant plusieurs étapes : Analyser et évaluer l'intérêt et les vulnérabilités de l'entité visée.

Il s'agit de sélectionner les cibles sur lesquelles l'effort militaire sera concentré en cohérence avec les objectifs de planification et les ressources (humaines, techniques) disponibles afin de monter les opérations nécessaires pour traiter les cibles dans un souci d'efficience et de limitation au maximum des effets négatifs (dommages collatéraux).

[10] http://www.opex360.com/2018/02/06/face-a-linflation-donnees-direction-renseignement-militaire-mise-lintelligence-artificielle/ du 6 février 2018

Trois champs d'affrontement immatériels ont été identifiés jusqu'à présent : le champ des perceptions, l'environnement électromagnétique et le cyberespace.

Ils interagissent avec les milieux physiques et comme ces derniers, sont eux-mêmes étroitement interdépendants. Du fait du développement rapide et de la diffusion mondiale des Technologies de l'Information et de la Communication (TIC), la maîtrise du champ des perceptions et du cyberespace est devenue un enjeu militaire déterminant[11].

L'exécution des actions de ciblage passe par la mise en œuvre de moyens d'acquisition et de traitement des cibles. L'optimisation du processus de ciblage et du traitement des cibles requiert la maîtrise des modalités techniques de mise en œuvre de systèmes interopérables et d'équipements interconnectés.

Le ciblage large spectre est une fonction stratégique positionnée au niveau du Centre de Planification et de Conduite des Opérations (CPCO) de l'état-major des armées. Le ciblage large spectre reprend les principes du ciblage mais sur des domaines élargis et des champs matériels comme immatériels. Les cibles sont définies après une étude systémique large qui repose sur tous les domaines structurants de l'adversaire. La principale difficulté du ciblage large spectre est d'équilibrer et de synchroniser les actions et l'engagement de cibles variées sur des champs liés de façons variables. Le ciblage large spectre doit pouvoir anticiper les effets synchronisés sur des objectifs militaires et interministériels. L'exemple des bombardements massif sur les populations est parlant. Les tapis de bombes sur les populations civiles allemandes dès 1943 ont prouvé leur inefficacité. En effet, pour avoir l'effet de sape sur le moral ennemi, les frappes doivent être couplées à une action d'influence. La question est de savoir comment et quand synchroniser les actions létales et coercitives avec la délivrance des messages choisis. Aucun système de simulation ne permet à ce jour d'anticiper de façon éclairée les effets des actions coordonnées.

L'analyse prédictive pourrait proposer des scénarios crédibles et des probabilités d'occurrences qui seraient précieuses pour les décideurs.

B. Des réponses à des besoins identifiés

L'organisation des états-majors opératifs et tactiques répond à une logique initiale construite en tuyau d'orgue : la planification, la conduite, le renseignement, la

[11] Doctrine interarmées 3.9 (DIA-3.9) « Ciblage »

logistique, les ressources humaines et financières, le retour d'expérience, les communications, etc...

Il paraît pourtant évident que ces domaines sont liés y compris sur un tempo temps réel. Afin de cadrer la charge de travail et les objectifs de chaque domaine, il est impensable d'ouvrir totalement les cellules et de procéder à un décloisonnement absolu au risque d'aboutir à une logique d'emploi du temps saturé dans lequel s'enchaineraient des suites de réunions sans fin dont une partie seulement intéresse chaque intervenant.

Afin de satisfaire le besoin de connaissance transverse et partagé, le calcul permanent d'une image « tactique-opérative-ressource » paraît nécessaire. Il ne semble pas raisonnable d'imaginer un système fonctionnant d'emblée et capable de répondre à ce besoin global dès la première minute de mise en fonctionnement. Cependant, il paraît beaucoup plus pertinent la construction progressive d'un système répondant d'abord aux besoins de chaque tuyau et qui puisse répondre ensuite à des besoins transverses en cumulant des corrélations. C'est ce que les technologies prédictives permettent. Il s'agira ensuite, lorsque le système sera éprouvé, de rechercher plus en profondeur les conclusions de niveau opérativo-stratégiques.

Dans un premier temps, chaque domaine (logistique, conduite, renseignement, etc.) nécessite d'avoir une vision proche du temps réel de la situation et des options disponibles. Ce que l'OTAN appelle la « *picture* ». Des systèmes de calcul reliés à des capteurs pertinents peuvent y répondre assez rapidement. Les logiciels de dialogue de gestion, de suivi tactique, de suivi logistique ou de cartographie tactique par exemple répondent à ce besoin de « *picture* ». Ces logiciels doivent pouvoir trier les informations pertinentes pour chaque domaine afin de répondre aux besoins cruciaux à temps. Dans le domaine de la conduite, des informations relatives à des conditions d'ouverture du feu qui seraient réunies sont prioritaires à des informations relatives à l'environnement social des cibles. Cependant, les informations « annexes » ne doivent pas être effacées. Elles peuvent avoir une signification à un niveau supérieur.

Pour être efficaces et rentables dans chaque tuyau et chaque domaine, les logiciels doivent fonctionner comme des lanceurs d'alertes. Ainsi, une contrainte logistique majeure peut être anticipée, une localisation estimée actionnable d'une cible peut être établie, des contraintes financières évitées.

Dans un deuxième temps, une image transverse peut être construite à partir des conclusions partielles des systèmes répondant à chaque tuyau.

Les états-majors fonctionnent comme cela. Les technologies analytiques et le *big data* peuvent apporter une plus-value dans l'identification de situations particulières en anticipation. Cela s'avère particulièrement intéressant lorsque la complexité de l'environnement et des facteurs concourants ne permet pas à des analystes humains de produire cette anticipation ou à tout le moins, pas en temps contraint.

Ainsi, Les informations initialement filtrées aux niveaux tactiques et par domaines peuvent avoir de la pertinence au niveau transverse. Cela comprend l'analyse des signaux faibles. Certaines informations sur l'adversaire, inconnues jusque-là, peuvent être identifiées sans être décelées ou démontrées ni même prises en compte. Par exemple, la présence d'un proxy d'un individu cible à haute valeur ajoutée (*high value target individual*) peut déclencher toute une série de signaux faibles que le système intègrera. Par corrélations successives, le système produira par l'analyse de ces signaux l'identification de la présence du proxy alors que cette information n'est pas connue ni démontrée ni même soupçonnable via une analyse logique. Une fois de plus, l'importance et l'architecture des corrélations est primordial. Ainsi, le fonctionnement des systèmes dans la durée et sur un nombre important d'événement est un facteur d'efficacité et de pertinence de l'analyse prédictive. Cela répond tout particulièrement aux besoins liés aux opérations dites de stabilisation. En effet, ces opérations sont le plus souvent longues et caractérisées par un ennemi fugace, insaisissable et son action sur la population n'est pas connue.

C. Des contraintes structurelles. (Dont le travail collaboratif multidisciplinaire « élargi »)

Les données analysées dans le cadre du ciblage large spectre doivent provenir de tous les domaines interarmées et interministériels. Elles doivent recouvrir les aspects économiques, culturels, militaires, politiques ou juridiques. Les structures étatiques actuelles demeurent cloisonnées. Il n'existe pas à ce jour de serveur commun ni même de passerelle d'exploitation des données brutes.

Le partage des données au sein du ministère des armées fait également face à des cloisonnements. Les armées, direction et services ne dialoguent que ponctuellement. La séparation entre les fonctions organiques, opérationnelles et de maintenance

demeure un obstacle. Techniquement, les différents niveaux de confidentialité des informations et des données interdisent leurs échanges sur un réseau commun.

De plus, la qualité des réseaux varie en fonction des contraintes physiques rencontrées dans les déploiements. Le partage des données et des informations dans un contexte multinational interallié reste un vrai sujet complexe. La confiance accordée à chaque partenaire diffère *a priori.*

Ainsi, l'approche multidisciplinaire relative au ciblage large spectre ne peut s'appuyer sur un réseau exhaustif ni même sur une vaste base de données (*data lake* ou magnadata). La construction d'une architecture permettant une analyse collaborative semble difficile Cela nécessitera probablement plusieurs étapes progressives n'engageant qu'un petit nombre d'acteurs à chaque fois.

Enfin, les décideurs doivent avoir accès aux conclusions et aux simulations produites par les systèmes numériques sans recourir à des connaissances techniques poussées. De manière générale, la convivialité et la souplesse des interfaces doit être particulièrement soignée.

L'ensemble des acteurs doit avoir conscience que les systèmes prédictifs ne font que fournir des hypothèses. En aucun cas, ils ne comprennent un phénomène et surtout, ils ne portent pas la responsabilité d'une décision. Les différents échelons décisionnels doivent donc clairement définir leurs places dans un système en partie automatisé.

D. Des contraintes de saisie, de disponibilité, de partage et de validité des données

Le nombre des données produites et disponibles explosent. Le phénomène du *Big data* est souvent décrit comme répondant à la règle des 3V : volume, vitesse et variété. Le volume décrit la quantité de données générées. La vitesse décrit la fréquence à laquelle les données sont générées, capturées et partagées. La variété des données traduit la prolifération de types de données. Ces 3V explosent du fait de l'évolution continue des technologies liées au numérique. Le volume et la vitesse des données disponibles révèlent un paradoxe, elles constituent à la fois une opportunité d'établir une description plus précise du monde réel ainsi qu'une difficulté majeure et croissante à traiter ces données. La variété est devenue une problématique car la prolifération de types de données provenant de sources comme les médias sociaux, les interactions *Machine to Machine* et les terminaux mobiles, crée une très grande diversité au-delà

des données transactionnelles traditionnelles. Les données ne s'inscrivent plus dans des structures nettes[12].

Afin d'obtenir une traduction numériquement exploitable, le monde doit être considéré comme un ensemble d'évènements. Chaque évènement se caractérise par une information propre à laquelle sont associées un nombre plus ou moins important de métadonnées. Par exemple, l'information d'une explosion dans une ville peut avoir comme métadonnées associées les coordonnées géographiques de l'explosion, l'altitude, le groupe date-heure de l'explosion, le groupe date-heure du relais de l'information, les références de la source, etc… Ces métadonnées comportent en général des informations précieuses d'un poids numérique faible (quelques bits). Il est impératif de concevoir que dans un univers militaire, les données nécessitent d'être filtrées.

Plusieurs niveaux (de sécurité, de protection, de partage) qui dépendent des relations diplomatiques ou d'estimation de fiabilité caractérisent chaque information. Ces différents niveaux animent les réseaux et se traduisent par la mise en place de filtres. Cette problématique de filtre est réelle car lorsqu'un évènement est capturé, il est soumis à des processus de filtrage plus ou moins complexes. Le temps de latence du processus doit rester minime pour ne pas perdre trop d'information en cas de crash d'une des composantes du système.

Il est également primordial de disposer de points de contrôle réguliers qui évite d'avoir à traiter des évènements passés et donc obsolètes. Le flot des évènements à traiter comprend généralement entre 10000 et 1000000 événements par seconde. Passé 24heures, ce nombre peut atteindre 100 milliards d'événements par jour pour un volume de 30 Terabits à stocker en moyenne. Cela explique qu'aucune solution simple n'ait encore été trouvée.

Afin de faire face à ces contraintes temporelles, plusieurs approches sont possibles pour obtenir une version digitalisée acceptable du monde. La première consiste à utiliser des puissances de calculs croissantes pour traiter de façon exhaustive chaque événement et limiter le flux de données. Une autre approche parait particulièrement pertinente : afin d'obtenir une version digitalisée exploitable du monde, il s'agit de considérer une approximation qui ne tient compte que des informations *a priori*

[12] journaldunet.com/solutions/expert/51696/les-3-v-du-big-data

pertinentes. Ainsi : plusieurs principes doivent être suivis et acceptés: une version digitalisée peut être efficace sans être complète ; cette version est spécialisée sur un domaine (le renseignement militaire par exemple) ; elle est basée sur des données qui doivent être récoltées avec une intention et un objectif spécifiques ; elle est basée sur des algorithmes classant du plus simple au plus complexe qui utilisent les corrélations pour reconstruire une version digitalisée du monde ressemblante et enfin la version digitalisée minimale du monde doit être partagée afin que les procédés analytiques puissent proposer une meilleure interprétation de la réalité.

E. Les cycles du ciblage

L'Allied Joint Publication 3.9 (AJP-3.9) « ALLIED JOINT DOCTRINE FOR JOINT TARGETING » définit deux types distincts de « *targeting* », c'est-à-dire de ciblage : le ciblage planifié (*deliberate targeting*) et le ciblage accéléré (*dynamic targeting*). La doctrine interarmées 3.9 (DIA-3.9) « Ciblage » ajoute un troisième type, le *Time Sensitive Targeting* ou ciblage prioritaire immédiat. Le ciblage délibéré traite de cibles planifiées connues pour être présentes dans un environnement donné. Ce type ciblage caractérise les situations où les données relatives à chaque cible sont suffisamment détaillées pour que le cycle du ciblage interarmées puisse être planifié et conduit. Cela permet d'assigner des moyens et des procédés particuliers à l'engagement (cinétique ou non) de ces cibles. Le ciblage dynamique traite de cibles connues pour exister dans une zone d'opération mais ne sont pas détectées, localisées ou actionnables dans le cadre du processus « délibéré ». Dans ce cas, le cycle du ciblage interarmées classique ne peut être exécuté proprement pour engager les cibles concernées.

Le cycle du ciblage interarmées se caractérise par six phases : phase 1, analyse des objectifs; phase 2, la phase de *target development* qui se traduit par identification et approbation des cibles; phase 3, choix des capacités ; phase 4, répartition par composantes et modalités d'exécution ; phase 5, l'exécution des missions et phase 6 : l'évaluation des effets.

L'exécution et la conduite des missions de ciblage de la phase 5 répondent d'un nouveau cycle particulier constitué de sept étapes : le cycle « F2T2E2A » pour *find, fix, track, target, engage, exploit* et *assess.* Ce cycle est utilisé en particulier pour le ciblage dynamique. Les opérations spéciales utilisent un autre cycle inspiré de la doctrine américaine qui se rapproche du F2T2E2A. Il s'agit du cycle F3EAD *find, fix,*

finish, exploit, analyse, disseminate. Ce cycle est particulièrement adapté aux opérations de contre-terrorisme.

Il en sera fait référence dans le chapitre 2.7 cas particulier des opérations spéciales.

Le ciblage est donc caractérisé par un cycle décisionnel. Ce cycle s'étoffe avec la complexité des crises. Le nombre de paramètres et de facteurs à analyser s'accroit. La charge de travail associée s'alourdit et le cycle de décision se ralentit. Les capacités de calculs des systèmes informatiques, plus rapides par essence que l'analyse humaine, associées aux algorithmes de corrélation permettent de retrouver un tempo acceptable. Mieux, cet outil prédictif permet d'intégrer en temps réels des éléments nombreux et variés. Il permet également de joindre aux analyses des informations corrélatives.

Avec le temps et la rotation du cycle de ciblage, la précision des systèmes augmente et permet d'atteindre des niveaux inaccessibles à la seule analyse humaine pour un temps de production restreint. Cette analyse prédictive permet également de simuler plusieurs options, toujours en utilisant le principe des corrélations. Ces simulations offrent aux chefs militaires opératifs et stratégiques des options en avance de phase. En utilisant l'analogie avec une partie d'échec, plusieurs options et coups d'avances peuvent ainsi être proposés. Cet atout majeur dans la résolution des crises intéresse en particulier le Sous-Chef Opération de l'État-major des Armées (SCOPS). Ainsi le général de corps de Saint Quentin a confié au CPOIA (Centre de Préparation aux Opérations InterArmées) le pilotage d'un projet baptisé C2 LAB.

F. Des simulations stratégiques

Différentes sociétés industrielles proposent des moyens de simulation. La base industrielle et technologique de la défense française présente un nombre important d'entreprises à même de développer des systèmes numériques capables de fournir des simulations de niveaux tactiques. Des sociétés qui investissent le secteur défense comme THALES, SAFRAN, DASSAULT, MBDA ou EADS possèdent la technologie et les ressources suffisantes pour développer des simulations pertinentes sur des problématiques techniques qui permettent à la fois de travailler la formation et l'instruction comme d'anticiper des risques potentiels.

Cependant, à ce jour, aucun système de simulation ne répond aux questions stratégiques et n'effectue des simulations poussées jusqu'à ce niveau.

La croissance des flux de données, les capacités de traitement rapides ainsi que des algorithmes prédictifs pertinent laissent envisager à termes que des simulateurs stratégiques large spectre puissent être élaborés.

G. Cas particulier des opérations spéciales

Les opérations spéciales offrent au pouvoir politique des options originales pour traiter des objectifs particuliers. Elles jouent un rôle primordial dans la lutte contre le terrorisme. Le mode de fonctionnement peu habituel des forces spéciales et la portée stratégique de leurs actions imposent un contrôle au plus haut niveau. Dans les opérations spéciales, le domaine tactique se mêle bien souvent aux questions et impératifs du niveau stratégique. Les opérations spéciales mettent en œuvre des états-majors plus resserrés. Ces états-majors permettent des actions plus directes, plus intégrées mais également plus transverses et plus interarmées que les opérations conventionnelles. Ce cadre particulier est le milieu adapté à l'expérimentation de l'analyse prédictive de de l'exploitation du big data dans des missions de ciblage.

Ainsi, une opération spéciale visant à engager ou à agir sur des objectifs à haute valeur ajoutée et qui exploitera toutes les potentialités des calculateurs aura un tempo opérationnel particulièrement rapide. En effet, les calculateurs permettent d'obtenir des informations pertinentes en temps réel ou en temps réflexe.

IV – Perspectives et recommandations

A. Les différences entre les logiques de guerre et les logiques commerciales

Les analogies avec le monde du marketing ou de la finance doivent être tempérées. En effet, la logique du commerce ou de l'économie consiste à obtenir un maximum de rentabilité en exploitant un terrain commun aux compétiteurs. Là où la logique de guerre correspond à une dialectique des volontés. Ainsi, le ciblage commercial ou électoral consiste à prendre plus que son concurrent sur un terrain commun, là où le ciblage militaire consiste à prendre sur son adversaire. Le ciblage économique consiste à choisir ses efforts sur un segment de population indépendant a priori du concurrent. Là où le ciblage militaire consiste à choisir ses efforts sur l'adversaire lui-même.

Le ciblage militaire doit donc considérer que l'adversaire évoluera et cherchera également à produire des effets sur ce qui est pour lui son adversaire. Parmi les différences fondamentales, on notera le cas où devancer un compétiteur dans le domaine économique aura pour limitation le fait de ne pas provoquer une diminution de rentabilité. Alors que dans le cas d'un affrontement militaire, diminuer la puissance de chaque protagoniste est acceptable si cela conduit à la victoire.

Les simulateurs prédictifs ont donc un intérêt pour prédire les réactions de l'adversaire mais une campagne militaire doit considérer l'enchainement de réaction de toutes les parties. Le ciblage constitue un domaine qui permet de conceptualiser ses propres actions.

Afin de définir les meilleurs effets sur les champs physiques et immatériels, le recours à la simulation permet d'étudier plusieurs enchainements possibles. Le recours à l'analyse prédictive présente donc un intérêt certain dans le domaine du ciblage. Sa pertinence reste encore à étudier. En quoi l'utilisation de super calculateurs peut-elle constituer un gain ?

Pour répondre à cette question, il s'agit d'observer deux problèmes : y-a-t-il suffisamment de données concernant un adversaire et son environnement pour mener une analyse ? Cette masse de donnée peut-elle être analysée ?

B. L'investissement en ressource humaine

Les armées ont besoin d'investir en ressource humaine pour pouvoir mettre en œuvre efficacement les technologies numériques. Cet investissement s'avère aujourd'hui assez lourd. Les degrés de qualification sont élevés, la ressource est rare. Les salaires proposés par le secteur privé se rapprochent des grilles de soldes des officiers supérieurs dans des postes à responsabilités. De plus, les sciences de la donnée exigent pour être efficace de l'innovation permanente et une connaissance de plus en plus précise de l'environnement. Le principe de mobilité et de mutation régulière est un obstacle actuellement à l'exploitation à termes des compétences des spécialistes de la donnée.

Ainsi, il parait nécessaire de disposer de personnel formé a minima dans la donnée mais connaissant parfaitement le domaine militaire pour lequel il utilise les algorithmes. Pour compléter cela, des noyaux d'experts de la donnée doivent être mis en place. L'investissement est donc nécessaire en termes de *data analysts*, dans tous

les domaines d'expertise militaire, un nombre moindre de *data scientists* et enfin, un noyau de *developpers* et de *data scientists* pour corriger les erreurs techniques et pousser plus en avant la puissance des algorithmes.

L'intelligence artificielle ne supprime pas des emplois pour l'homme, elle déplace le besoin vers des techniciens de la donnée et des réseaux. Bien souvent ce besoin est plus important en termes de ressource humaine. L'intérêt réside dans l'efficacité ainsi obtenue. Ce principe ne doit pas être ignoré dans les forces armées. Les domaines du renseignement et de la maintenance en particulier ont un intérêt critique aujourd'hui d'investir dans les compétences liées aux sciences de la donnée. Le domaine du ciblage restera l'échelon ultime dans lequel le personnel qualifié et compétent sera nécessaire. Il est cependant primordial. En effet, le ciblage large spectre est probablement le domaine le plus transverse du niveau stratégique pour l'exploitation des données.

C. L'importance croissante de la question de l'intelligence artificielle

« En juin 2016, le centre de recherche de l'*US Air Force* (AFRL – *US Air Force Research Laboratory*) et l'entreprise PSIBERNETIX ont réalisé une expérience ayant donné un aperçu de ce que pourrait être la façon de faire la guerre dans les décennies à venir. Ainsi, il a été mis au point une intelligence artificielle qui, appelée ALPHA, a été imaginée pour mener une patrouille d'avions de chasse en situation de combat» rappelle Laurent Lagneau, journaliste spécialisé dans la Défense.

L'intelligence artificielle ouvre de nouvelles perspectives dans le domaine militaire en général. L'exemple type est l'essaim de drones. Leur développement est symptomatique des inquiétudes que suscite les robots armés. L'intelligence artificielle permettra d'automatiser la prise de décision et l'analyse. La question de la responsabilité des décisions se pose alors. Dans les règles d'engagement adoptées par les forces armées en opération. L'automatisation des décisions devra être cadrée.

Aussi, Jean-Yves Le Drian, alors ministre de la Défense a rappelé en 2017 « l'intelligence artificielle est un élément de notre souveraineté nationale[13] ». D'après les propos rapportés par l'Usine Nouvelle et Acteurs Publics, il estime qu'il s'agit même d'une « troisième révolution stratégique, dite troisième offset » après « l'hypervélocité et la lutte sous-marine. » Et c'est d'ailleurs la raison pour laquelle, « nos alliés », et en

[13] Jean-Yves Le Drian, Colloque "Intelligence artificielle : des libertés individuelles à la sécurité nationale", Assemblée nationale le 14 février 2017.

particulier les États-Unis, l'ont placée « au cœur de la stratégie de sauvegarde de leur défense et de leur souveraineté. »

« Il s'agit de créer une troisième rupture technologique, après la dissuasion nucléaire et l'explosion des technologies de l'information et du numérique, pour garantir la supériorité et la sécurité américaine. Cette révolution potentielle sert de cadre et d'aiguillon à une politique d'investissement audacieuse », a expliqué M. Le Drian.

Il a également évoqué ce que serait le successeur du Rafale. « Il faut compter sur l'intelligence des plateformes, leur capacité à se reconfigurer, à dialoguer entre elles, et parfois être simplement sacrifiées, ce qui n'a de sens qu'avec des plateformes pilotées à distance, capable elles-mêmes d'attaques saturantes si nécessaires ».[14]

« La robotisation associée aux capacités d'intelligence artificielle s'imposera inéluctablement sur le champ de bataille en raison de ses nombreux atouts. La fonction létale de ces robots ne sera qu'une option additionnelle à des objets relevant de technologies duales. Il est donc difficile d'imaginer que certains belligérants ne seront pas tentés de se doter de tels systèmes d'armes. Le scénario de rupture qui pourrait être envisagé à l'échéance de 2030 serait donc le recours à des systèmes entièrement autonomes dotés de capacités létales (SALA). Ce risque serait d'autant plus grand que les SALA, dotés d'une intelligence artificielle, utiliseraient leur capacité d'autoapprentissage pour s'éloigner des règles initiales d'ouverture du feu. Cette autonomisation serait aussi très dangereuse compte tenu du risque de déprogrammation et de reprogrammation des robots par des opérations cybernétiques adverses. Cette autonomisation, qui aurait pour corollaire un désengagement de l'être humain du champ de bataille, transformerait assurément la physionomie de la guerre qui n'aurait plus comme limites que les capacités de robots, démultipliées par rapport à celles des êtres humains » [15].

D. L'investissement technologique

L'augmentation constante de la vitesse des microprocesseurs et l'augmentation massive de la ressource en données permet d'envisager la nécessité croissante de disposer d'algorithmes adaptés à l'exploitation des systèmes. Ces algorithmes ne seront

[14] http://www.opex360.com/2017/02/17/m-le-drian-fait-de-lintelligence-artificielle-enjeu-strategique-pour-la-defense/ du 17 février 2017

[15] CHOCS FUTURS, Étude prospective à l'horizon 2030 : impacts des transformations et ruptures technologiques sur notre environnement stratégique et de sécurité, SECRÉTARIAT GÉNÉRAL DE LA DÉFENSE ET DE LA SÉCURITÉ NATIONALE, p198

vraisemblablement pas universels et devront être réétudiés, révisés ou réécrits en fonctions des modèles désirés. Ainsi chaque simulateur, chaque nouvelle situation exigera une adaptation même minime de ces logarithmes. Il ne s'agit pas seulement de systèmes apprenants se mettant à jour mais aussi d'une réorientation du système en fonction des nouveaux produits désirés. Ainsi, il apparaît nécessaire d'envisager un moyen de disposer d'une ressource humaine à même de répondre à ces exigences.

Elle doit manifestement se définir à plusieurs niveaux : depuis les experts en programmation jusqu'aux simples opérateurs de mise à jour ainsi que le personnel combattant ou identifié aujourd'hui comme opérationnel à même de définir les produits requis.

La technologie numérique constituera naturellement un nouveau domaine de puissance. Le rythme auquel la popularité d'une application pour smartphone chasse l'autre, donne probablement le tempo d'adaptabilité auquel ce domaine aura à faire face.

En partant de ces présuppositions, la solution du DGA LAB semble être une première pierre dans la construction d'un outil en mesure de relever ces défis technologiques. A nouveau, le caractère éphémère doit être considéré et l'organisation à même de répondre à la performance de l'outil technologique numérique doit être définie en fonction de ce critère prioritaire.

Il s'agit maintenant de distinguer le challenge assez général de l'intelligence artificielle et des calculateurs de celui de l'analyse prédictive à proprement parlé. Les modèles prédictifs faisant appel aux technologies du *deep data* et aux capacités heuristiques et donc apprenants devront être confrontés à un maximum de situations réelles ou se rapprochant de la réalité. A nouveau, le fait que la France soit régulièrement engagée dans des conflits complexes laisse envisager un fort potentiel de développement et de pertinence de ces systèmes. Ceci est d'autant plus vrai que les théâtres potentiels sont listés parmi des zones d'intérêt définies par le niveau stratégique et donc a priori connue et regorgeant d'événements antérieurs pertinents pour une analyse prédictive fine et précise.

Il semble assez abordable de disposer d'outils de faible portée, éphémères et pour des applications précises. Cependant l'accès à des systèmes prédictifs de niveau stratégique ou même opératif semble complexe. En effet, l'analyse opérative fonctionne par

synthèses des différents niveaux subalternes successifs, or une analyse prédictive pertinente fonctionnerait et fonctionne déjà dans d'autres domaines de façon itérative et analytique. C'est à dire que ces systèmes étudient chaque facteur aussi minime et granulaire soit-il. Ces systèmes seraient donc dans l'idéal, capable d'anticiper les phénomènes dit d'effet papillon, totalement hors de portée d'une analyse humaine générale classique fonctionnant par synthèses des niveaux inférieurs.

Ces systèmes semblent difficiles à mettre en œuvre étant donné le nombre de connexions à des sources de données et de facteurs, de modèles variés et différents. Cette sorte de super moteur de recherche et d'analyse « Google » de la vraie vie se heurterait fatalement à un immense travail de normalisation. La difficulté normative a déjà suffisamment été éprouvée par les armées et la longueur de temps d'aboutissement des grands programmes d'équipement témoignent d'une première contrainte structurelle.

Il semble donc raisonnable d'envisager à nouveau des systèmes d'analyse segmentés en niveaux, étudiant les conclusions d'autres systèmes de niveaux inférieurs dans un premier temps. Cependant, une totale mise en connexion doit rester envisageable car à termes, la faisabilité technique paraît inéluctable.

A l'image de la DARPA (*Defense advanced research projects agency*), un pôle français de développement technologique semble nécessaire. La DARPA offre la possibilité à des entités innovantes de développer leur concept. Plusieurs chefs de projet utilisent la collaboration participative Elle s'appuie sur un « écosystème d'innovation » qui comporte des universitaires, des sociétés de recherche et développement, des partenaires gouvernementaux, tous en permanence orientés vers les armées.

L'agence revendique environs une centaine de directeurs de programmes qui dirigent autours de 250 projets en tout. Elle dispose d'un budget officiel de 3,17 milliards de dollars.

Plusieurs ruptures technologiques ont été établies par la DARPA. On distingue pour les capacités militaires : les munitions guidées, les technologies furtives et pour les technologies à forts intérêts civils : les systèmes de reconnaissance vocale, le GPS (*Global Positioning System*) ou bien Internet. Il semble intéressant de souligner que plusieurs projets de la DARPA portent sur l'exploitation des données et les systèmes prédictifs. Ces travaux ne se limitent pas à la recherche fondamentale mais participent

également à résoudre des problématiques de la ressource et en particulier des ressources humaines. Ainsi, après avoir identifié le manque d'expert dans le domaine des sciences de la donnée. La DARPA travaille sur un programme qui permettrait à termes de construire des modèles empiriques de façon automatique ou partiellement automatique. Ce programme permettrait de développer des systèmes prédictifs pour des utilisateurs sans connaissance particulière dans les sciences de la donnée.

Le nom attribué à ce programme est le *Data-Driven Discovery of Models* (D3M). Son but est donc de permettre à des non-spécialistes de créer des modèles prédictifs empiriques complexes en automatisant une large partie des procédés de création des modèles.

Le déficit en expert serait alors compensé par des systèmes accessibles à des planificateurs militaires plus généralistes mais également aux spécialistes du renseignement qui pourraient chercher plus finement les paramètres pertinents. Ainsi, dans le domaine du ciblage, les différents critères pourraient être définis plus précisément. On imagine sans difficulté des simulations qui mettraient en valeur l'intérêt de tel ou tel critère.

Ainsi, une structure en lien avec les technopoles régionaux pourrait idéalement reproduire à une échelle accessible aux ressources françaises un modèle de DARPA « à la française ».

Il semble crucial de valoriser la convivialité des programmes et donc de travailler une interface homme-machine efficace. En effet, la disponibilité des utilisateurs pour apprendre à utiliser ces systèmes est assez faible. Les planificateurs et officiers du renseignement ont avant tout besoin d'un fond solide en expérience militaire opérationnelle voire en fonctionnement structurel.

E. Les faiblesses des systèmes

Les systèmes faisant appel aux technologies informatiques et des réseaux se heurtent aujourd'hui aux menaces cyber. Même si on ne s'intéresse qu'à la capacité se calcul, à l'utilisation des algorithmes et aux traitements des données, le fait que ces données proviennent de différentes sources, n'exclut pas le risque d'interférence ni d'intrusion voire d'introduction d'un malware. Il convient donc de définir une architecture qui ne fonctionne qu'avec des éléments sains et de limiter le dialogue avec des sources mal maitrisées. Les liaisons sécurisées doivent être privilégiées à l'image des liaisons de

données tactique (L11, L16, L22) ou bien cryptées. Les sources d'origines ouvertes doivent être contrôlées ou à défaut, les données utilisées doivent être formatées. Plusieurs couches de protection peuvent également constituer l'architecture d'ensemble.

Une autre faiblesse des systèmes liés au *big data* réside dans les besoins matériels concrets nécessaires pour faire fonctionner l'ensemble des réseaux, des calculateurs, des moyens de stockage et des capteurs. Ainsi, les systèmes informatisés demandent une ressource énergétique importante. Des actions sur les moyens d'approvisionnement énergétique risque de dégrader fortement le système en imposant au mieux des coupures qui nécessitent un relai d'un autre moyen d'approvisionnement en énergie et au pire, une mise en sommeil d'une large partie du système le laissant bancal et peu pertinent. Les coupures brèves ne constituent pas un risque important à courts termes mais si elles sont répétées, provoquent des retards dans le traitement des données qui perturbent l'établissement d'une *picture* ou d'une image en temps réel. De plus, les serveurs de données nécessitent pour fonctionner, des locaux adaptés, dont la température doit être régulée. L'espace nécessaire à l'installation des serveurs s'avère être aussi une contrainte. De plus, l'approvisionnement en énergie de ces serveurs est un enjeu majeur. Ces lieux de stockage demeurent des cibles de choix à protéger. Le rayonnement résultant du fonctionnement des serveurs et des systèmes annexes est particulièrement signant.

Enfin, comme tout réseau, les nœuds de communications et les voies d'accès des données peuvent constituer des faiblesses s'ils sont attaqués. Citons par exemple, les câbles de fibre optique, les terminaux de connexions, les antennes relais ou bien les centres d'émission et de réception. La fabrication technique de ces moyens informatique exige des connaissances particulières ainsi que des matériaux rares qui peuvent à terme poser des problèmes en cas de besoin de renouvellement.

V – Conclusion

Les technologies liées au phénomène baptisé *big data* sont en plein essor et constituent un facteur d'évolution considérable dans nombre de domaines. La tendance n'est pas à un ralentissement des progrès mais bien à une poursuite des évolutions, une augmentation des

performances et un élargissement des champs d'action. Le fait que les objets sont de plus en plus connectés, que le nombre d'objets connectés augmente, que les connections se multiplient et que les capacités de stockage progressent impose de s'intéresser au phénomène. L'accès à un grand nombre de données et l'amélioration constante du traitement de ces données a déjà permis des progrès dans les domaines de la logistique, du commerce en ligne et du renseignement.

En parallèle de ce phénomène, les développements des sciences de l'information automatisée, de l'algorithmie et des statistiques ont considérablement amélioré les performances dans des domaines importants comme la sécurité intérieure, l'analyse électorale et surtout la finance. La science des algorithmes liée aux puissances de calcul de l'informatique a prouvé son efficacité implacable en fournissant progressivement des systèmes capables de vaincre les humains dans tous les jeux de stratégie quels qu'ils soient.

Ces deux domaines se lient maintenant pour ouvrir de nouvelles dimensions dans de nombreuses disciplines. Les systèmes résultant de l'exploitation de ces nouvelles technologies offrent des progrès considérables. Ces progrès se manifestent dans la vitesse de traitement des problèmes, dans la précision des réponses apportées mais également dans la pertinence des simulations proposées et les applications lanceuses d'alerte associées à ces systèmes de simulation et d'anticipation.

Les performances ainsi offertes ont imposé la mise en place de ces systèmes dans la plupart des domaines concurrentiels exploitant l'information. Cependant, l'exploitation de ces capacités numériques exige également le recours à des spécialistes, à des formations particulières ainsi qu'un investissement constant dans la recherche et le développement des systèmes.

La Défense et les armées ne sont pas exclues du phénomène et révèlent des besoins particuliers qui méritent de poursuivre ou redoubler d'intérêt pour ces technologies.

Le domaine du ciblage en particulier ainsi que son enveloppe, le ciblage large spectre sont directement concernés par la pertinence des systèmes prédictifs. Il apparait de plus en plus probable qu'ils puissent jouer un rôle accru dans le domaine du ciblage.

En effet, ce dernier est caractérisé par un cycle, dont le tempo peut varier mais dans lequel tout gain de temps est précieux. La masse et la variété des données accessibles augment, compliquant ainsi le travail d'exploitation. L'emploi de calculateurs s'avère nécessaire pour conserver un rythme de cycle pertinent. L'utilisation grandissante du numérique se constate de

façon de plus en plus fréquente essentiellement à des niveaux tactiques mais aussi dans le domaine plus large du renseignement. La mise en œuvre des systèmes prédictifs associés à l'explosion du *big data* pourrait potentiellement révolutionner l'approche stratégique du ciblage sous tous ses aspects. Pourtant, de nombreuses contraintes tant au niveau structurel qu'en termes de ressources doivent être surmontées pour obtenir un système efficace, rentable et pertinent.

Il permettrait de fournir des simulations stratégiques précieuses, d'identifier des faiblesses masquées chez l'ennemi et de fournir des options claires au pouvoir politique. Dans cette entreprise, le cas des opérations spéciales constitue un laboratoire potentiel pour l'expérimentation large d'un système prédictif. Il demeure que l'organisation générale des armées et des états-majors opérationnels ne facilitent pas la mise en œuvre et l'exploitation d'un système numérique en temps réel et transverse. La complexité des réseaux de communication, à commencer par leur nombre, leur variété et les barrières les séparant rend difficile la mise en œuvre d'emblée d'un système. La constitution d'un logiciel et d'un réseau stratégique ne pourra donc se faire que progressivement et par briques.

Les outils de l'analyse prédictive et d'exploitation du *big data* exigent pour être opérants et efficaces plusieurs investissements. Un investissement en termes de ressources humaines est nécessaire pour concevoir, exploiter, modifier et faire évoluer les systèmes numériques. Un vivier de spécialistes est nécessaire. Le personnel qualifié est rare et sa rémunération reste un sujet. De plus, l'architecture de la carrière militaire de ce personnel doit être adaptée. Un second investissement doit être consenti. Il s'agit de l'investissement relatif à la mise en place de réseaux et de logiciels évolutifs et modulaires. Enfin, il est nécessaire d'anticiper la création de structures physiques capables d'accueillir des serveurs.

L'ensemble de cet investissement doit être rentabilisé. Pour cela, la pertinence du système à mettre en place doit être bien cernée. Si le ciblage commercial a été optimisé par l'exploitation du *big data* et que l'analyse prédictive constitue l'outil principal de la finance, ces logiques marchandes sont différentes de la gestion de crise ou de l'affrontement qui caractérise la fonction des forces armées. De même, l'intelligence artificielle utilisée par les systèmes numériques reste fondée sur les corrélations et le calcul anticipé. Ainsi, elle doit être adaptée à un environnement changeant avec des règles qui se modifient sans cesse. Le cas de figure du monde réel n'est pas celui du jeu de go.

Miser sur ces technologies s'avère être un challenge et constitue de nombreux risques. L'analyse prédictive ne constituera définitivement un *game changer* que lorsqu'elle aura

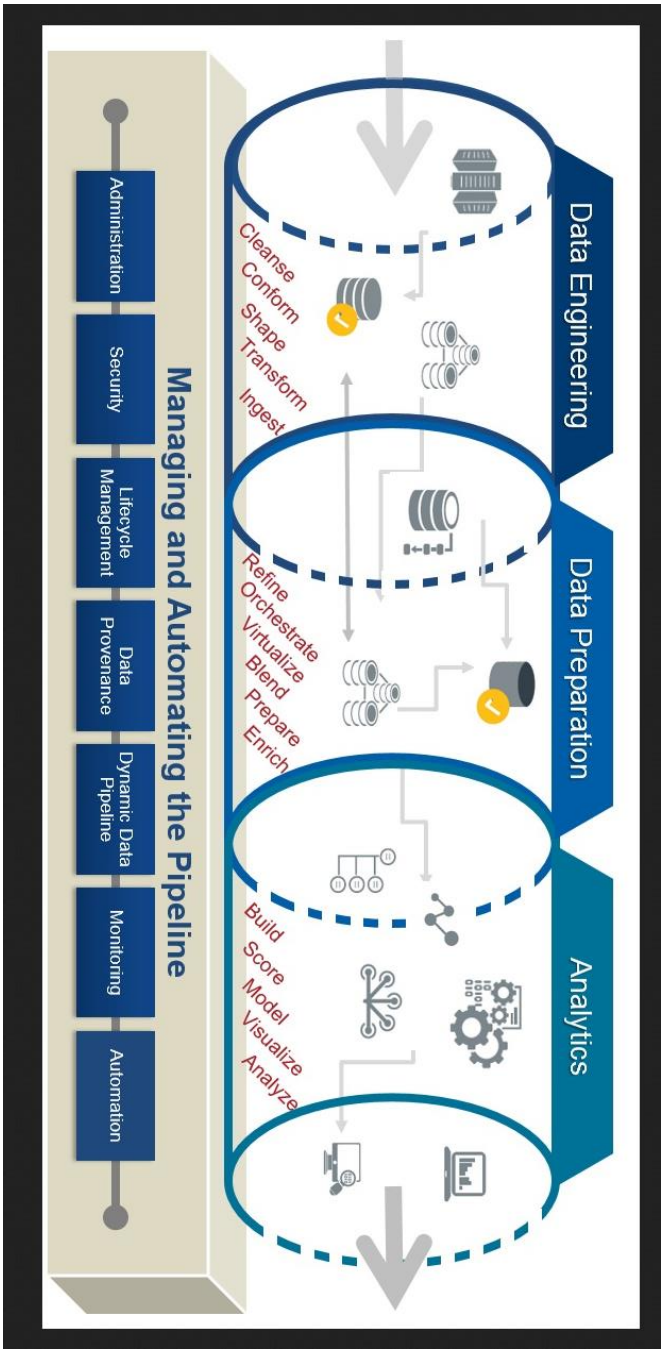
atteint le domaine du ciblage large spectre. Pour l'instant, son importance est plus floue. Dans l'ensemble des domaines ayant eu recours à ces logiciels, on observe essentiellement une optimisation statistique et un gain d'efficacité chiffré. Les systèmes fondés sur des corrélations statistiques offrent ainsi une optimisation comparable à ces chiffres. L'efficacité non négligeable de ces systèmes doit être mesurée au regard de leur rentabilité. Une optimisation de 20 ou 30% ne mérite pas parfois l'investissement consenti.

Pour autant, il n'est pas certain que l'institution militaire ait vraiment le choix. Les géants du numérique, les GAFA, s'imposent comme de nouveaux pouvoirs en situation de quasi-monopole du fait de leur avance technologique. A cette image, il est facile de percevoir qu'un décrochage concurrentiel est possible voire probable. Il parait de plus évident, que la multiplication constante des capteurs ainsi que de leurs capacités va nécessiter des investissements. Et, à l'image des jeux de stratégie et du jeu de go, il deviendra un jour impossible à un acteur de l'emporter sans avoir recours à un système prédictif.

En France, la question éthique, voire philosophique retient plus l'attention que la recherche technique ou l'approfondissement des technologies et des possibilités potentielles. La question de l'intelligence artificielle est essentiellement traitée par des sociologues. Les œuvres littéraires depuis Frankenstein jusqu'à Les Robots d'Isaac Asimov ont plus marqué les esprits que les plans d'urbanisation fonctionnels et leurs déclinaisons en lignes de code. Pourtant, Cédric Villani, député de la cinquième circonscription de l'Essonne et en charge depuis septembre d'une mission portant sur l'intelligence artificielle souligne combien les inquiétudes liées au domaine ne s'appuient que sur des phantasmes inspirés par la science-fiction. La vraie crainte subsiste dans le risque de décrochage. La vitesse d'évolution du concept -dans les possibilités techniques comme dans l'engouement qu'il suscite- font qu'entre le début de la rédaction de ce mémoire et sa conclusion, d'importantes nouveautés ont vu le jour, provoquant en partie la remise en question de sujets traités. C'est symptomatique du risque que pourrait courir la France si elle rate le rendez-vous de la révolution numérique. Il s'agit donc de ne pas négliger ces disciplines.

Enfin, l'investissement continu dans l'analyse prédictive permettra de concevoir des systèmes à l'intelligence artificielle supérieure aux systèmes adverses et ainsi de contrer leurs attaques potentielles et en particulier les attaques à caractère terroriste. Une des meilleures réponses à la menace des systèmes autonomes armés n'est pas de les interdire sur le plan du droit international mais de développer des systèmes qui rendent ces armes obsolètes.

ANNEXES


Data Engineering
Cleanse
Conform
Shape
Transform
Ingest
Data Preparation
Refine
Orchestrate
Virtualize
Blend
Prepare
Enrich
Analytics
Build
Score
Model
Visualize
Analyze
Managing and Automating the Pipeline
Administration
Security
Lifecycle Management
Data Provenance
Dynamic Data Pipeline
Monitoring
Automation

Le cycle de vie des données au sein du Big Data
1- Recueil
2 - Stockage
3 - Traitement
4 - Utilisation
Dossier Client
Données officielles
Géolocalisation
Profils des réseaux sociaux
Historiques de navigation
Algorithmes Requêtes
Textmining Framework Hadoop
Modélisation Visualisation
Relation Client
Efficacité Opérationnelle
Stratégie
Source : Exeis Conseil

BIBLIOGRAPHIE

Entretiens libres:

- Colonel Martin, chef de la section ciblage large spectre du Centre de planification et de conduite des opérations ;
- Frédéric Larbre, architecte des systèmes d'information chez ING direct.

Ouvrages consultés:

- Marz, N.,Warren, J. « Big Data Principles and Best Practices of Scalable Realtime data Systems » Manning Publications Co 2012 ;
- Babak Akhgar, Gregory B. Saathoff, Hamid R. Arabnia, Richard Hill, Andrew Staniforth, Petra Saskia Bayerl. « Application of Big Data for National Security ». Butterworth-Heinemann 17 février 2015 ;
- Thomas Cambrai. « L'intelligence artificielle expliquée : Comment les algorithmes et le Deep Learning dominent le monde ! ». Thomas Cambrai 20 novembre 2017.

Documents de doctrine :

- Doctrine interarmées 3.5 (DIA-3.5) « Opérations spéciales » ;
- Doctrine interarmées 3.9 (DIA-3.9) « Ciblage » ;
- Allied Joint Doctrine for Joint Targeting (AJP-3.9).

Articles :

- https://www.futura-sciences.com/tech/big-data/ de 2017 ;
- http://www.le big data.fr de 2017;
- http://www.lemagit.fr de 2017;
- http://www.blogdumoderateur.com/metiers-data-regionsjob 2017;
- http:/www.mrmint.fr 2017 ;
- https://www.blogdumoderateur.com/metiers-data-regionsjob du 12/12/2017;
- http://www.previoo.com, octobre 2016 ;

- https://news.sap.com/france/2017/09/21/cas-dutilisation-de-lintelligence-artificielle-au-profit-de-la-chaine-logistique/ du 12 mars 2018 ;
- http://www.opex360.com/2018/02/06/face-a-linflation-donnees-direction-renseignement-militaire-mise-lintelligence-artificielle/ du 6 février 2018 ;
- journaldunet.com/solutions/expert/51696/les-3-v-du-big-data ;
- http://www.opex360.com/2017/02/17/m-le-drian-fait-de-lintelligence-artificielle-enjeu-strategique-pour-la-defense/ du 17 février 2017 ;
- Pearsall, Beth. « Predictive Policing : The Future of Law Enforcement. » National Institute of Justice Journal 266 (2010).

Autre :

- CHOCS FUTURS, Étude prospective à l'horizon 2030 : impacts des transformations et ruptures technologiques sur notre environnement stratégique et de sécurité, SECRÉTARIAT GÉNÉRAL DE LA DÉFENSE ET DE LA SÉCURITÉ NATIONALE, p198