



**HAL**  
open science

# “ Autonomie ” dans les systèmes d’armes : questionnements sémantiques, techniques, éthiques

Catherine Tessier

## ► To cite this version:

Catherine Tessier. “ Autonomie ” dans les systèmes d’armes : questionnements sémantiques, techniques, éthiques. Les enjeux de l’autonomie des systèmes d’armes létaux, Mar 2022, Paris, France. hal-03740293

**HAL Id: hal-03740293**

**<https://hal.science/hal-03740293>**

Submitted on 29 Jul 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L’archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d’enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

## « Autonomie » dans les systèmes d'armes : questionnements sémantiques, techniques, éthiques

Catherine Tessier

Référente intégrité scientifique et éthique de la recherche de l'ONERA

Membre du Comité d'éthique de la défense et du Comité national pilote d'éthique du numérique

« Une machine autonome est une machine automatique dans son milieu. » (Laumond, 2021)<sup>1</sup>

On lit et on entend qu'un robot, un système d'armes, un véhicule, est qualifié d'« autonome », voire de « pleinement autonome », ou bien de « partiellement » ou « semi autonome », ou encore, comme « intégrant de l'autonomie » (Comité d'éthique de la défense, 2021)<sup>2</sup>. De même de tels systèmes sont susceptibles d'intégrer de l'« intelligence artificielle », voire sont qualifiés d'« intelligents ». Un tel vocabulaire masque des enjeux essentiels qu'il faut étudier et dont il faut débattre : de quelles entités parle-t-on exactement ? que cherche-t-on à automatiser et pourquoi ? quelles techniques sont employées ? quelles tensions se présentent ?

### 1 Cessons de parler d'« autonomie » et d'« intelligence artificielle »<sup>3</sup>

Les machines automatisées ne sont ni une fatalité ni un phénomène qui émergerait plus ou moins naturellement. Ce sont en effet des personnes – chercheurs, ingénieurs, utilisateurs – qui ont décidé que certaines fonctions d'un objet seraient automatisées. Ainsi un robot est « *une machine qui bouge et dont les mouvements sont commandés par un ordinateur* » (Laumond, 2012)<sup>4</sup> : un robot est équipé de moyens de calcul permettant de déterminer des actions à réaliser, avec leurs modalités et paramètres, en fonction d'un état particulier de l'environnement qui est estimé à partir de données recueillies par des capteurs puis traitées et mises en correspondance avec des références préétablies. Cette machine est conçue pour répondre à un usage particulier – le nombre et la nature des actions qu'elle peut effectuer sont limitées – dans un milieu donné.

Dans l'expression « système d'armes autonome », il est important de noter que l'adjectif « autonome » est donc employé selon l'acception de la robotique : un robot « autonome » est une machine mobile dotée de capacités programmées de perception et de calcul d'actions qu'elle effectue dans un milieu déterminé, sans intervention extérieure, ces actions étant spécifiées dans un cahier des charges précis. Or l'acception courante du mot « autonomie », lorsqu'il est employé à propos d'êtres vivants et particulièrement d'êtres humains, est la faculté de déterminer ses propres objectifs et de choisir librement ses actions. Il se crée alors une confusion entre les acceptions, source de malentendus, d'erreurs, de craintes ou d'attentes infondées, y compris au sein des instances internationales de négociation : une arme « autonome » serait imprévisible par nature, prendrait d'« elle-même » des

---

<sup>1</sup> Jean-Paul Laumond, *Robotique : l'intelligence artificielle face au réel*, Mythes et machines, Académie des sciences, novembre 2021, <https://www.youtube.com/watch?v=sTU70egicvM>

<sup>2</sup> Comité d'éthique de la défense, *Avis sur l'intégration de l'autonomie dans les systèmes d'armes létaux*, avril 2021, [https://www.defense.gouv.fr/salle-de-presse/communiqués/communiqué\\_le-comite-d-ethique-de-la-defense-publie-son-rapport-sur-l-integration-de-l-autonomie-des-systemes-d-armes-letaux](https://www.defense.gouv.fr/salle-de-presse/communiqués/communiqué_le-comite-d-ethique-de-la-defense-publie-son-rapport-sur-l-integration-de-l-autonomie-des-systemes-d-armes-letaux)

<sup>3</sup> Voir aussi : Catherine Tessier, *Parler du numérique et de son éthique : un questionnement... éthique*, in Comité national pilote d'éthique du numérique, *Pour une éthique du numérique*, coordonné par Éric Germain, Claude Kirchner, Catherine Tessier, PUF 2022

<sup>4</sup> Jean-Paul Laumond, *La robotique : une récidive d'Héphaïstos*, Leçon inaugurale prononcée au Collège de France, 2012, <https://books.openedition.org/cdf/518>

« initiatives » ou « déciderait » de ses propres objectifs (Comité national pilote d'éthique du numérique, 2021)<sup>5</sup>.

D'autre part, l'expression « intelligence artificielle » est redevenue à la mode, et dans des usages laissant penser que l'on parle d'une entité ayant le même statut moral qu'un être vivant. Ainsi « l'intelligence artificielle prend des décisions », « on apprend à une intelligence artificielle », « une intelligence artificielle doit être digne de confiance », etc. De plus, bien que l'intelligence artificielle soit un ensemble de techniques très variées (AFIA)<sup>6</sup>, l'acception largement répandue actuellement est limitée à celle qui recouvre les techniques d'apprentissage numérique : « intelligence artificielle » est considéré comme un synonyme de « système à base d'apprentissage ». *A contrario*, certains discours considèrent que l'intelligence artificielle inclut des techniques telles que la théorie de la commande, l'analyse numérique, etc. Il conviendrait donc d'abandonner l'effet de mode et de nommer précisément les outils scientifiques dont on parle.

En outre, un amalgame peut être fait entre « autonomie » et « intelligence artificielle » – pris dans son acception réduite : ainsi un système d'armes « autonome » serait un système d'armes « apprenant », et de surcroît continuant d'« apprendre » au cours de sa mission sans contrôle approprié.

Les confusions ainsi propagées amènent à débattre de questions dépourvues de sens, à nourrir des fantasmes et surtout à ne pas s'atteler aux questions scientifiques et éthiques de fond. Par exemple, la tentative de comparaison de l'intelligence et de l'intelligence artificielle est une erreur fondée sur l'emploi du mot « intelligence » sous deux statuts différents : l'intelligence artificielle n'est pas une intelligence qui serait artificielle, mais le nom d'une discipline scientifique – que l'on devrait plus précisément et objectivement remplacer par « informatique » ou « numérique » ; en outre nous concevons (grâce à notre intelligence) des fonctions automatisées pour justement effectuer des tâches qui ne requièrent pas d'intelligence (Steels, 2021)<sup>7</sup>. Comme le résume Luc Steels<sup>8</sup> : « *nous devrions peut-être cesser de parler d'intelligence artificielle* » ; et cesser de parler de l'« autonomie » des machines.

## 2 Automatiser : quoi et pourquoi ?

### 2.1 Raisonner selon les fonctions

Considérer un objet dans sa globalité pour raisonner sur son automatisation n'est pas pertinent, puisque selon les fonctions de cet objet qui sont automatisées et les milieux et contextes dans lesquels elles sont mises en œuvre, les enjeux scientifiques et éthiques sont différents. Par exemple une arme létale peut être équipée uniquement de fonctions automatisées de planification et d'exécution des déplacements dans l'espace ; elle peut aussi être équipée de fonctions automatisées d'identification, de classification, d'interception, d'engagement<sup>9</sup>. Il s'agit donc de poser la question suivante : quelles fonctions précisément seront automatisées dans la machine considérée, c'est-à-dire programmées sur un ordinateur – souvent embarqué dans la machine ?

---

<sup>5</sup> Comité national pilote d'éthique du numérique (CNPEN), *Avis n°2 : Le véhicule « autonome » : enjeux d'éthique*, avril 2021, <https://www.ccne-ethique.fr/fr/actualites/cnpen-le-vehicule-autonome-enjeux-dethique>

<sup>6</sup> Association française pour l'Intelligence Artificielle, Mots-clés de l'AFIA, <https://afia.asso.fr/domaines-de-lia/lafia-en-25-mots-cle/>

<sup>7</sup> Luc Steels, *Defining AI for purposes of ethics and legal regulations*, AI4EU Workshop: The Culture of Trustworthy AI. Public debate, education, practical learning, September 2021, <https://www.youtube.com/watch?v=1B2dsOZcxfA> (from 17:00)

<sup>8</sup> *Ibid.*

<sup>9</sup> Comité d'éthique de la défense, *Avis sur l'intégration de l'autonomie dans les systèmes d'armes létaux*, avril 2021, [https://www.defense.gouv.fr/salle-de-presse/communiques/communiqu%C3%A9\\_le-comite-d-ethique-de-la-defense-publie-son-rapport-sur-l-integration-de-l-autonomie-des-systemes-d-armes-letaux](https://www.defense.gouv.fr/salle-de-presse/communiques/communiqu%C3%A9_le-comite-d-ethique-de-la-defense-publie-son-rapport-sur-l-integration-de-l-autonomie-des-systemes-d-armes-letaux)

## 2.2 Pourquoi automatiser une fonction ?

Les raisons qui poussent à automatiser des fonctions, en particulier dans un système d'armes, sont multiples – voir aussi l'avis du Comité d'éthique de la défense<sup>10</sup> :

- le problème posé n'est pas directement appréhendable par un être humain : par exemple il requiert des calculs complexes, l'exploitation de grandes quantités de données, ou est par nature combinatoire ;
- la fonction automatisée sera plus « performante » que la même fonction réalisée par des opérateurs humains : vitesse de réaction ou d'exécution plus grande, volume d'événements traités plus important ;
- l'automatisation assurera la permanence de la fonction, dans le temps et dans l'espace ;
- la fonction automatisée sera plus économique que la même fonction réalisée par des opérateurs humains ;
- la fonction automatisée permettra d'éloigner les opérateurs humains de zones dangereuses ou hostiles ;
- la fonction automatisée sera plus sûre que la même fonction réalisée par des opérateurs humains et en particulier, permettra de pallier les imprécisions ou les erreurs humaines.

L'automatisation de fonctions doit donc répondre à des besoins bien identifiés vis-à-vis de leur rôle, de leur réalisation et de leurs résultats. Ces besoins doivent être exprimés précisément dans un cahier des charges.

Il convient également d'évaluer la part des considérations suivantes – reprises et adaptées de (Murphy and Moats, 2021)<sup>11</sup> :

- L'automatisation des fonctions répond-elle effectivement à un besoin identifié et exprimé par les futurs utilisateurs, ou bien est-elle poussée par le marché de l'innovation ? Poser cette question peut permettre d'éviter de répondre à un besoin inexistant, de créer de toutes pièces un nouveau besoin, ou de concevoir des fonctions automatisées inadaptées, qui seront mal, peu, voire pas, utilisées.
- « Quelque chose [des fonctions automatisées] est mieux que rien [pas de fonctions automatisées] ». Cependant, les inconvénients présentés par les fonctions automatisées ont-ils été explicités et examinés au regard des avantages recherchés ? Par exemple le milieu dans lequel les fonctions vont être mises en œuvre est-il susceptible de présenter des situations qui n'ont pas été anticipées lors de la conception (les « inconnus inconnus ») et que ces fonctions ne pourront pas traiter de manière appropriée ? Les dysfonctionnements et modes dégradés des fonctions automatisées ne sont-ils pas pires que « rien » ?
- « Les autres le font ». Quels sont les arguments, les critères, les valeurs, l'éthique, les expériences, des « autres » ? Comment faire en sorte d'éviter le décrochage technologique<sup>12</sup> tout en continuant à respecter nos principes et valeurs ? Comment nos positionnements éthiques sont-ils susceptibles de

---

<sup>10</sup> *Ibid.*

<sup>11</sup> Robin R. Murphy and Jason B. Moats, *Responsible robotics for disaster response*. IROS 2021 Workshop "Building and Evaluating Ethical Robotic Systems", October 2021, <https://ers-workshop.com/pdf/MMers21.pdf>

<sup>12</sup> Voir principe P4, Comité d'éthique de la défense, *ibid.*

varier en fonction des possibilités techniques, de l'évolution de la société, de la nature des menaces ?

### 3 Automatiser : comment ?

#### 3.1 Quelle modélisation ?

Automatiser une fonction (considérée ici comme non triviale, c'est-à-dire ne se limitant pas au déclenchement d'une réaction sur stimulus) consiste à programmer cette fonction sur un ordinateur. Il s'agit de modéliser, donc de formaliser avec des concepts et outils informatiques ou mathématiques, les objets ou situations à identifier, les actions à réaliser, les critères de calcul de ces actions, les incertitudes, etc.

Le premier constat est qu'il existe des situations ou concepts qu'il est difficile, voire impossible, de modéliser, parce qu'il n'en existe pas de définition stricte, ou bien parce qu'ils ne peuvent être regroupés dans des classes définies par des caractéristiques communes. Ainsi le principe de proportionnalité du Droit international humanitaire n'est pas défini précisément, justement parce c'est le jugement humain qui, dans une situation particulière, va estimer qu'une action respecte ou non ce principe. De même, si les tous les terrains d'aviation présentent globalement des similarités, ce n'est pas le cas pour des situations tactiques terrestres, par nature originales et imprévisibles. Il y a donc une différence fondamentale entre une fonction de reconnaissance, que l'on pourra automatiser à partir de références modélisées<sup>13</sup>, et l'appréciation d'une situation qui, si elle inclut des aspects de reconnaissance, implique un faisceau d'éléments que seul l'être humain est en mesure d'appréhender.

Le second constat est que la modélisation est souvent une approximation de la réalité : un modèle numérique ne peut saisir toute la complexité, toutes les nuances, du monde réel. La démarche de modélisation implique donc des simplifications, des choix, des omissions (Kuipers, 2021)<sup>14</sup>, tout en étant suffisamment représentative pour l'objectif poursuivi. La fonction programmée ne peut ainsi effectuer des calculs qu'à partir de ces modèles. Ce qui n'a pas été modélisé, ni envisagé comme un risque identifié de sortie des modèles (« inconnus connus » – *known unknowns*), constitue des « inconnus inconnus » – *unknown unknowns*<sup>15</sup>. Les conditions d'usage de la fonction doivent être strictement définies<sup>16</sup>, et il est certainement illusoire d'envisager l'automatisation de fonctions d'ouverture du feu dans des milieux où il existe une grande incertitude sur ce qui est susceptible de se passer.

En outre, des questions doivent être posées au sujet des choix de conception et des méthodes employées :

- a-t-on suffisamment de recul sur les techniques et le matériel, ou bien sont-ils nouveaux et peu éprouvés ?<sup>17</sup>
- l'automatisation de la fonction résulte-t-elle d'une transposition d'un domaine (civil) à un autre (militaire) ?<sup>18</sup> cette transposition est-elle pertinente ?

---

<sup>13</sup> Notons cependant qu'une fonction automatisée de reconnaissance (d'objets par exemple) ne « reconnaît » rien ; elle effectue des calculs, des corrélations.

<sup>14</sup> Benjamin Kuipers, *What Can We Count On? Evaluating Ethical Robots*. IROS 2021 Workshop "Building and Evaluating Ethical Robotic Systems", October 2021, <https://ers-workshop.com/>

<sup>15</sup> *Ibid.*

<sup>16</sup> Voir recommandation R10, Comité d'éthique de la défense, *ibid.*

<sup>17</sup> Murphy and Moats, *ibid.*

<sup>18</sup> *Ibid.*

- les choix de modélisation et de programmation (architecture, paramètres, seuils, fonctions de coûts, etc.) sont-ils explicités et analysés ?<sup>19</sup>

### 3.2 Comprendre ou ne pas comprendre

Les utilisateurs des fonctions automatisées doivent comprendre ce que font ces fonctions, et être informés de ce qu'elles ne font pas, de ce qu'elles ne remplacent pas ; ils doivent également être conscients que ces fonctions sont susceptibles d'altérer leur jugement<sup>20</sup>. Cependant doivent-ils comprendre comment sont programmées ces fonctions ?

Les fonctions automatisées devraient être conçues de manière à se « rapprocher » de l'intuition humaine, au moins par les résultats qu'elles fournissent : ceux-ci doivent pouvoir être interprétés par l'utilisateur. Néanmoins une situation opérationnelle ou le Droit international humanitaire n'ont pas de sens pour une machine ou un logiciel<sup>21</sup>. Comment donc produire un sens à partir de quelque chose qui n'en a pas ? En outre, l'intégration dans les fonctions automatisées de traitements qui s'éloignent des capacités humaines (par exemple, l'affectation d'une signature spectrale à chaque pixel d'une image) est susceptible de créer des artéfacts (c'est-à-dire d'éléments qui n'ont pas d'existence dans la réalité) lors de la phase de mise en forme pour présentation à l'opérateur.

Sauf pour les concepteurs et programmeurs des fonctions automatisées, les processus mis en œuvre dans ces fonctions sont difficilement appréhendables, voire complètement opaques (qui sait comment exactement fonctionne son *smartphone* ?). En revanche, l'utilisateur doit pouvoir cerner ce que va faire ou ce qu'est susceptible de faire la fonction – le service rendu – sans nécessairement comprendre comment elle le fait.

### 3.3 Vérification et validation

Lorsqu'on ne peut prouver parfaitement une fonction automatisée complexe (par exemple, en démontrant formellement qu'elle respecte strictement le cahier des charges), comment en construire les critères de vérification, de validation, de qualification, voire de certification ? Cette question implique un certain nombre de défis scientifiques :

- inclure les utilisateurs et leurs actions dans le processus de qualification ;
- apporter les preuves de la conformité de la réalisation des tâches par la machine par rapport aux spécifications, quelles que soient les évolutions de l'environnement ;
- évaluer les risques d'interprétation des capacités de la machine au-delà de ses capacités réelles (les tests ne sont pas la réalité)<sup>22</sup> ;
- pour les fonctions à base d'apprentissage numérique, requalifier en cycle « court » les mises à jour effectuées à partir de nouvelles données ;
- évaluer les impacts des défaillances sur la mission ; recenser les possibilités de reconfiguration ou de réparation ainsi que leurs disponibilités et leurs coûts ; qualifier les conditions de désactivation d'urgence<sup>23</sup> des fonctions automatisées.

---

<sup>19</sup> Mélanie Gornet, *Enjeux d'éthiques liés aux algorithmes d'apprentissage machine et mise en œuvre technique de recommandations*. Rapport de stage ISAE-SUPAERO-SciencesPo-CNPEN, septembre 2021

<sup>20</sup> Voir recommandation R4, Comité d'éthique de la défense, *ibid.*

<sup>21</sup> Notons que cette phrase ne devrait pas être écrite afin de ne pas personnifier la machine ou le logiciel.

<sup>22</sup> Luc Steels, September 2021, *ibid.*

<sup>23</sup> Voir recommandation R17, Comité d'éthique de la défense, *ibid.*

## 4 Conclusion

Imaginons un système de ciblage automatique fondé sur la reconnaissance faciale ou posturale. S'il est montré que ce système ne fonctionne pas bien, c'est-à-dire que des individus, dans certaines situations, ne sont pas « reconnus » ou sont associés à une identité qui n'est pas la leur, le système ne pourra être utilisé. Il s'agit en effet d'un système critique, c'est-à-dire tel qu'un mauvais fonctionnement entraîne des conséquences catastrophiques.

En revanche, on l'a vu plus haut, montrer que ce système fonctionne parfaitement est en général impossible à *démontrer*. Et que signifierait d'ailleurs un fonctionnement *parfait* ? Serait-ce un fonctionnement modélisant fidèlement ce que déciderait un être humain ? Cette question se heurte en particulier à deux écueils :

- la difficulté de définir l'être humain, ou du moins les décisions humaines, de « référence », pour construire le modèle ;
- la nature approximative d'un modèle, quelle que soit sa sophistication.

Ainsi au-delà de la tension entre ce qui est faisable et ce qui est souhaitable, la vigilance éthique consiste à toujours garder à l'esprit les limites intrinsèques de l'automatisation de fonctions, en particulier dans un système d'armes, et plus généralement dans tout système critique.