



## Trabajo Fin de Grado

# ÉTICA DEL EMPLEO DE SISTEMAS DE ARMAS AUTÓNOMAS Y DRONES

Autor

Arturo Aurelio Arias Gasco

Director/es

Director académico: Dra. Dña. Silvia Vicente Oliva  
Director militar: Tte. D. Aurelio Vidal Fernández Diez

Centro Universitario de la Defensa-Academia General Militar  
2022





## Agradecimientos

En primer lugar, me gustaría agradecer a mi directora académica, la Dra. Silvia Vicente, por su paciencia, disponibilidad permanente e interés. Sus consejos y experiencia me han ayudado enormemente en la realización de este proyecto.

Por otro lado, me gustaría agradecer a mi director militar el teniente D. Aurelio Fernández, por su profesionalidad y las enseñanzas que escapan a cualquier libro de valores. Estoy seguro que su ejemplo me guiará en mi futuro profesional, así como me ha guiado durante la realización de este trabajo.

Adicionalmente, quiero dar las gracias a todo el personal de la Compañía de Apoyo, especialmente al capitán Rafael Sánchez Mata y al teniente Gonzalo de la Torre Díaz por su acogida en Burgos durante mi periodo de prácticas.

No me olvido de mis profesores, compañeros y amigos que me han acompañado durante este proyecto y los exigentes años de formación previos. Sin ellos, no me habría convertido en el militar, ingeniero y persona que estoy orgulloso de ser.

Por último, me gustaría dar las gracias a mis padres, por su apoyo incondicional, espero que todo el esfuerzo realizado durante estos años les haga sentirse orgullosos, os quiero.





# RESUMEN

El rápido avance tecnológico en distintos ámbitos, especialmente en el militar, genera importantes transformaciones en tácticas, técnicas y procedimientos. La situación cambiante en la que los ejércitos están constantemente inmersos muestra la necesidad de realizar un análisis ante posibles consecuencias de los cambios tecnológicos y su impacto en la sociedad.

Los avances desarrollados en el área de la automatización, con el objetivo de conseguir la plena autonomía de determinados sistemas de gran atractivo en la industria militar de defensa debería abrir un debate ético acerca del uso de los mismos con acciones coordinadas como consecuencia.

La finalidad de este trabajo es dar a conocer el desarrollo que han protagonizado las armas autónomas y los drones militares, así como dar respuestas acerca de las implicaciones éticas derivadas del uso de estos sistemas y proponer una serie de medidas que permita regular su empleo en futuros conflictos.

Para ello se ha realizado un análisis de prospectiva mediante prototipos de ciencia ficción complementado con un estudio semicuantitativo del que han formado parte cinco expertos de diferentes materias, todas ellas relacionadas con las armas autónomas y drones militares.

Los resultados obtenidos mediante el empleo de la metodología de prospectiva complementado con el criterio de los expertos han permitido identificar los escenarios de aplicación de los sistemas, además de conceptos de importancia ética en relación a estos, como son la responsabilidad, el riesgo para la vida, el reconocimiento moral mutuo y el estatus moral.

Con el fin de mitigar el impacto negativo en relación al uso de estas armas, se han propuesto unas medidas tomando como base los conceptos identificados anteriormente. Estas medidas incluyen la no transferibilidad de responsabilidad hacia las máquinas, un análisis exhaustivo de los riesgos en todo el ciclo de desarrollo de estas, la cumplimentación de unos estrictos estándares de seguridad y la implementación de un código moral integrado en los algoritmos de decisión de los sistemas autónomos.

## Palabras clave

Ética, moral, inteligencia artificial, drones, autonomía.



# ABSTRACT

The rapid technological advance in different fields, especially in the military, generates important transformations in tactics, techniques and procedures. The changing situation in which armies are constantly immersed shows the need to carry out an analysis of the possible consequences of technological changes and their impact on society.

The advances developed in the area of automation, with the aim of achieving the full autonomy of certain highly attractive systems in the military industry, should open an ethical debate about their use with coordinated actions as a consequence.

The purpose of this work is to make known the development that autonomous weapons and military drones have led, as well as to give answers about the ethical implications derived from the use of these systems and to propose a series of measures that allow regulating their use in future conflicts.

To this end, a prospective analysis has been carried out using science fiction prototypes, complemented by a semi-quantitative study in which five experts from different subjects have taken part, all of them related to autonomous weapons and military drones.

The results obtained through the use of the prospective methodology complemented with the criteria of the experts have allowed to identify the application scenarios of the systems, as well as concepts of ethical importance in relation to these, such as responsibility, risk to life, mutual moral recognition and moral status.

In order to mitigate the negative impact related to the use of these weapons, some measures have been proposed based on the concepts identified above. These measures include the non-transferability of responsibility towards the machines, an exhaustive analysis of the risks throughout their development cycle, compliance with strict security standards and the implementation of a moral code integrated into the decision algorithms of the autonomous systems.

## KEYWORDS

Ethics, morality, artificial intelligence, drones, autonomy



## INDICE DE CONTENIDO

<i>Agradecimientos</i> .....	I
<b>RESUMEN</b> .....	III
<i>Palabras clave</i> .....	III
<b>ABSTRACT</b> .....	IV
KEYWORDS .....	IV
<b>INDICE DE FIGURAS</b> .....	VIII
<b>INDICE DE TABLAS</b> .....	IX
<b>ABREVIATURAS, SIGLAS Y ACRÓNIMOS</b> .....	X
<b>1 INTRODUCCIÓN</b> .....	1
<b>2 OBJETIVOS Y METODOLOGÍA</b> .....	3
<b>2.1 OBJETIVOS Y ALCANCE</b> .....	3
<b>2.2 METODOLOGÍA</b> .....	3
2.2.1 ANALISIS MEDIANTE PROTOTIPOS DE CIENCIA FICCIÓN .....	4
2.2.2 FICCIÓN DE DISEÑO ÉTICO .....	5
2.2.3 METODOLOGÍA PARA EL ANÁLISIS.....	6
2.2.4 CUESTIONARIO A EXPERTOS .....	8
<b>3 ANTECEDENTES Y MARCO TEÓRICO</b> .....	9
<b>3.1 ÉTICA Y MORAL</b> .....	9
<b>3.2 ¿QUÉ ES UN DRON MILITAR?</b> .....	10
<b>3.3 ¿QUÉ ES UN ARMA AUTÓNOMA?</b> .....	12
3.3.1 INTERACCIÓN HUMANA EN LA AUTONOMÍA DE LOS SISTEMAS DE ARMAS .....	14
<b>4 DESARROLLO: ANÁLISIS Y RESULTADOS</b> .....	17
<b>4.1 DESCRIPCIÓN DEL MUNDO, EXPLICACIÓN DEL PROTOTIPO Y PRESENTACIÓN DE LA CIENCIA.</b> .....	17
4.1.1 Caso 1: Stealth: La amenaza invisible (2005) .....	17



4.1.2 Caso 2: Yo, Robot (2004) .....	18
4.1.3 Caso 3: Terminator 2: el juicio final (1991) .....	18
4.1.4 Caso 4: Sueños de Robot, Isaac Asimov (1947) .....	19
<b>4.2 IDENTIFICAR EL PUNTO DE INFLEXIÓN CIENTÍFICA Y LAS INTERACCIONES ENTRE LA INNOVACIÓN Y LA SOCIEDAD. ....</b>	<b>19</b>
4.2.1 Caso 1: Stealth: La amenaza invisible (2005) .....	19
4.2.2 Caso 2: Yo, Robot (2004) .....	20
4.2.3 Caso 3: Terminator 2: el juicio final (1991) .....	20
4.2.4 Caso 4: Sueños de Robot, Isaac Asimov (1947) .....	20
<b>4.3 IDENTIFICAR EL PUNTO DE INFLEXIÓN HUMANA.....</b>	<b>20</b>
4.3.1 Caso 1: Stealth: La amenaza invisible (2005) .....	20
4.3.2 Caso 2: Yo, Robot (2004) .....	21
4.3.3 Caso 3: Terminator 2: el juicio final (1991) .....	21
4.3.4 Caso 4: Sueños de Robot, Isaac Asimov (1947) .....	21
<b>4.4 REFLEXIÓN ÉTICA Y LECCIONES APRENDIDAS.....</b>	<b>22</b>
4.4.1 Caso 1: Stealth: La amenaza invisible (2005) .....	22
4.4.2 Caso 2: Yo, Robot (2004) .....	22
4.4.3 Caso 3: Terminator 2: el juicio final (1991) .....	23
4.4.4 Caso 4: Sueños de Robot, Isaac Asimov (1947) .....	24
<b>4.5 ESTUDIO SEMICUANTITATIVO .....</b>	<b>27</b>
4.5.1 RESULTADOS EN LA ESCALA LIKERT .....	27
<b>5 CONCLUSIONES.....</b>	<b>32</b>
<b>5.1 LINEAS DE ACCIÓN FUTURA .....</b>	<b>33</b>
<b>6 REFERENCIAS BIBLIOGRÁFICAS .....</b>	<b>34</b>
<b>ANEXOS.....</b>	<b>37</b>
<b>Anexo I. SFP .....</b>	<b>38</b>
Paso 1: Elegir la ciencia y construir el mundo .....	38
Paso 2: Identificar el punto de inflexión científica .....	38
Paso 3: Considerar las interacciones recíprocas de la innovación y la sociedad.....	38
Paso 4: Identificar el punto de inflexión humana. .....	38



Paso 5: Reflexionar sobre lo aprendido .....	39
<b>Anexo II. Ficción de diseño ético .....</b>	<b>40</b>
Paso 1: Relato o explicación .....	40
Paso 2. Prototipo .....	40
Paso 3. Discurso.....	40
Paso 4. Posturas éticas .....	40
Paso 5. Desarrollo del carácter.....	40
Paso 6. Apelación retórica.....	40
<b>Anexo III. Tabla resumen evaluación del objetivo del diseño.....</b>	<b>42</b>
<b>Anexo IV. Cuestionario Administrativo a expertos.....</b>	<b>43</b>
<b>Anexo V. Listado de obras de ciencia ficción .....</b>	<b>44</b>
<b>Anexo VI. Modelo de los 6 Arquetipos.....</b>	<b>45</b>
1. Crecimiento y decadencia .....	45
2. Amenazas y nuevas esperanzas .....	45
3. Mundos baldíos .....	45
4. Poderes fácticos .....	45
5. Desorden .....	46
6. Inversión.....	46
<b>Anexo VII. Respuestas al cuestionario administrativo.....</b>	<b>47</b>
Primer encuestado .....	47
Segundo encuestado .....	48
Tercer encuestado.....	49
Cuarto encuestado .....	50
Quinto encuestado.....	50



# INDICE DE FIGURAS

Ilustración 1. Proceso de previsión de futuro, ilustrando las entradas dispares, la creación de futuros y el retroceso con acciones específicas (gates) y eventos incontrolables (flags). Fuente: elaboración propia a partir de Bennett and Johnson (2016).....	4
Ilustración 2. Diagrama de los 5 pasos del proceso SFP. Fuente: elaboración propia .....	5
Ilustración 3. MQ-1 Predator armado con misiles Hellfire. Fuente: Lt. Col. Leslie Pratt (2008) .	12
Ilustración 4. El bucle OODA. Fuente: elaboración propia. ....	14
Ilustración 5. Esquema del modo semiautónomo (Human In The Loop). Fuente: elaboración propia.....	15
Ilustración 6. Esquema del modo de autonomía supervisada (Human On The Loop). Fuente: elaboración propia.....	16
Ilustración 7. Esquema del modo plenamente autónomo (Human Out The Loop). Fuente: elaboración propia.....	16
Ilustración 8. Fotograma de la película “Stealth: La amenaza invisible” dirigida por Robert Cohen, en la parte superior se puede observar el prototipo UCAV autónomo al lado de una aeronave tripulada. Fuente: Gaumont Columbia Tristar Films. ....	19
Ilustración 9. Fotograma de la película “Terminator 2: el juicio final”, dirigida por James Cameron. Fuente: Carolco Pictures, Pacific Western; Lightstorm Entertainment, T2 Productions.....	21
Ilustración 10. Fotograma de la película “Yo, Robot”, dirigida por Alex Proyas. Fuente: Davis Entertainment Overbrook Entertainment.....	24
Ilustración 11 Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca del nivel de regulación. Fuente: elaboración propia.....	28
Ilustración 12. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de la influencia en la sociedad. Fuente: elaboración propia. ....	29
Ilustración 13. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de los efectos producidos. Fuente: elaboración propia.....	29
Ilustración 14. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de los riesgos. Fuente: elaboración propia. ....	30
Ilustración 15. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de la moralidad. Fuente: elaboración propia. ....	31
Ilustración 16. Representación de los 6 arquetipos. Fuente: Fergnani and Song (2020). ....	46



# INDICE DE TABLAS

Tabla 1. Ficha técnica del estudio semicuantitativo a realizar. Fuente: elaboración propia.....	8
Tabla 2. Tabla resumen del análisis SFP en la que se muestran las fases para cada uno de los casos estudiados. Fuente: elaboración propia.....	26
Tabla 3. Recopilación de las puntuaciones en la escala Likert otorgadas por los expertos a las preguntas del cuestionario. Fuente: elaboración propia. ....	27
Tabla 4. Matriz de ficción de diseño. Fuente: Jensen and Vistisen (2017). .....	41
Tabla 5. Evaluación del objetivo del diseño. Fuente: Baeza (2020). ....	42
Tabla 6. Perfil del primer encuestado. Fuente: elaboración propia .....	47
Tabla 7. Perfil del segundo encuestado. Fuente: elaboración propia.....	48
Tabla 8. Perfil del tercer encuestado. Fuente: elaboración propia .....	49
Tabla 9. Perfil del cuarto encuestado. Fuente: elaboración propia.....	50
Tabla 10. Perfil del quinto encuestado. Fuente: elaboración propia.....	50



# ABREVIATURAS, SIGLAS Y ACRÓNIMOS

ACS: Aegis Combat System

AT: Aerial Target

AWS: Aegis Weapon System

CIWIS: Close-in Weapon System

C-RAM: Counter Rocket, Artillery and Mortar

DFCS: Digital Flight Control System

F2T2EA: Find, Fix, Track, Target, Engage, Assess.

GPS: Global Positioning System

HALE: High Altitude Long Endurance

IA: Inteligencia Artificial

ICAO: International Civil Aviation Organization

IMINT: Imagery Intelligence

LAWS: Lethal Autonomous Weapon Systems

MALE: Medium Altitude Long Endurance

OODA: Observe, Orient, Decide, Act

RAE: Real Academia Española

RPA: Remotely Piloted Aircraft

RPAS: Remotely Piloted Aircraft System

RRII: Relaciones Internacionales

SFP: Science Fiction Prototyping.

SIGINT: Signals Intelligence

UA: Unmanned Aircraft

UCAV: Unmanned Combat Aerial Vehicle

UAS: Unmanned Aircraft Systems

UAV: Unmanned Aerial Vehicle



# 1 INTRODUCCIÓN

Los avances tecnológicos a lo largo de la historia, han provocado enormes cambios en la forma en la que las sociedades viven, se relacionan, se desplazan, se comunican y hasta en cómo se enfrentan. De la misma manera, el progreso tecnológico ha tenido gran influencia en las guerras, empezando como batallas terrestres, siguiendo con el avance de la tecnología naval y posteriormente al ámbito aeroespacial. Actualmente, existe un nuevo escenario de confrontación, el ciberespacio, que se encuentra en continua evolución y se considera crítico debido a la enorme dependencia que tiene la civilización presente, tanto económica, política como socialmente de él (Colesniuc, 2013).

Bien es sabido que la evolución sufrida en este ámbito en las últimas décadas ha sido de una gran magnitud, especialmente en las áreas de las telecomunicaciones e informática. Esta rápida evolución ha sido propiciada, en gran medida, por los conflictos bélicos, en los que la superioridad tecnológica fue desde la mitad del siglo XIX, el elemento decisivo (Aznar Fernández Montesinos, 2016). La carrera armamentística y tecnológica siempre ha constituido -desde el origen de los conflictos entre distintas civilizaciones-, la salvaguarda de la seguridad y de los intereses de las naciones.

Entre los nuevos tipos de tecnología que están transformando numerosos aspectos de la sociedad se encuentran la robótica, la inteligencia artificial y los drones (González Estévez, 2019). Este tipo de avances se tratan de tecnología dual, es decir, se desarrollan tanto en el ámbito civil como militar. Tanto es así, que en el sector privado se están realizando grandes avances, muchos de ellos cerca de superar a los de la industria militar. Sírvase de ejemplo el Robot da Vinci y sus numerosas aplicaciones en intervenciones quirúrgicas que requieren gran precisión o la aplicación del campo de la Inteligencia Artificial (IA) en el desarrollo de vehículos de conducción autónoma, entre otros (Villavicencio Mavrich, 2005).

Desde principios de siglo se ha producido un gran desarrollo en este ámbito con el objetivo de aumentar las capacidades de los ejércitos, lo que ha provocado una mayor automatización de las tareas, hasta tal punto que se ha conseguido alcanzar la autonomía en muchas de ellas. Esta delegación de tareas de los soldados a las máquinas transforma en gran medida la naturaleza de las nuevas guerras, en las que, además de reducir el daño en personal sufrido por los ejércitos, se desvanecen las responsabilidades en cuanto a daños colaterales y se incrementa el anonimato a la hora de intervenir.

Además, el espacio físico en el que se desarrollan las guerras de hoy es cada vez es más reducido, siendo este ocupado por el ciberespacio, eliminando la distinción entre guerra y no guerra; el conflicto entre estados sigue estando presente, pero otros actores entran en escena.

El concepto que define el estado actual es el de guerra de cuarta generación (Lind, 2004), que involucra multitud de aspectos útiles para desestabilizar o generar grandes consecuencias; desde un ataque informático hasta acciones en el ámbito de las finanzas, éstas generan un conflicto que no se corresponde con uno convencional pero cuyas consecuencias pueden tener incluso más impacto.

La nueva forma de hacer guerra unido al empleo de nuevas tecnologías, abre un debate en cuanto a los límites de lo que es o no es guerra y lo que es aún más importante, qué implicaciones éticas tiene el empleo de estos nuevos avances tecnológicos en los conflictos actuales y los que están por venir. En el contexto en el que se realiza este trabajo, la ética, según la Real Academia Española (RAE), hace referencia a: "parte de la filosofía que trata del bien y del fundamento de los valores". Es necesario aclarar que la ética no trata de unos valores objetivos e invariables, sino que estos difieren entre las distintas sociedades, de ahí la dificultad de establecer un código ético que regule las distintas acciones. Debido a esto, es de gran importancia aclarar ciertos



conceptos, con el fin de establecer un marco regulatorio que rija el uso de estos sistemas y que se anticipe a los posibles escenarios futuros.

En este marco, el trabajo se desarrollará utilizando información de expertos y como principal herramienta de análisis los Prototipos de Ciencia Ficción (SFP) complementada mediante una metodología de ficción de diseño ético, con la finalidad de dar respuesta a las cuestiones éticas derivadas de estos sistemas en un contexto futuro. Previamente, se expondrá el estado del arte y el marco teórico asociado a estas tecnologías y como herramienta secundaria de análisis se utilizarán distintas entrevistas con el fin de contrastar los resultados obtenidos. A partir de las conclusiones extraídas mediante la aplicación de este método, se buscará clarificar y dar solución a las cuestiones éticas del uso de las nuevas tecnologías emergentes en el ámbito militar, como son los drones o el uso de la inteligencia artificial en armas autónomas.



# 2 OBJETIVOS Y METODOLOGÍA

## 2.1 OBJETIVOS Y ALCANCE

El objetivo general que busca este trabajo es analizar las implicaciones éticas que tienen el uso de las nuevas tecnologías autónomas y de drones en el ámbito militar, así como en los posibles conflictos derivados de su uso. Este documento presenta una recopilación acerca del estado del arte de estas tecnologías y las posibles evoluciones que sufrirán de cara al futuro. En cuanto a los objetivos específicos se desarrollan los siguientes:

- Conocer que es un dron militar y un sistema de armas autónomo.
- Comprender los conceptos de ética y de autonomía, así como los distintos grados de automatización.
- Analizar el contexto futuro en el que operarán estas tecnologías.
- Identificar las implicaciones éticas que tienen el empleo de estos sistemas.
- Proponer medidas para permitir, regular o prohibir su uso basándose en un análisis ético previo.

El alcance del trabajo comprende el empleo de tecnologías autónomas y drones en el ámbito militar, particularmente en un contexto futuro en el que estas se encuentren notablemente desarrolladas, hasta tal punto que puedan entrar en conflicto con las decisiones tomadas por humanos. En cuanto a la realización del análisis, se ha desarrollado una metodología basada en el empleo de SFP y ficción de diseño ético, éste se va a limitar a las situaciones futuras que se van a describir en los diferentes escenarios de análisis.

## 2.2 METODOLOGÍA

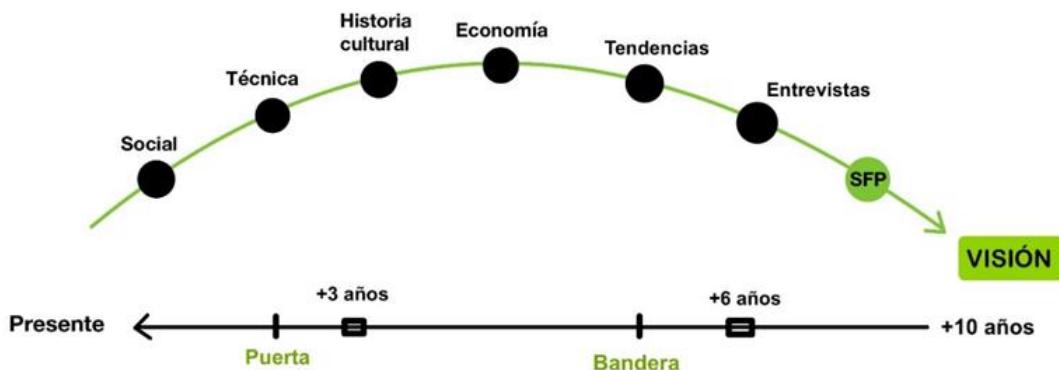
Conforme a los objetivos presentados anteriormente, la metodología del trabajo se basa en herramientas cualitativas. Este se compone de una descripción de los sistemas de armas autónomos y drones y aclaración de conceptos relativos al tema del trabajo.

Mediante el empleo de prospectiva (*foresight*), se pueden generar visiones viables para el futuro, combinando distintas áreas tales como ciencias sociales, investigación técnica, economía, análisis de tendencias, historia cultural, entrevistas con expertos y ciencia ficción. La prospectiva permite visualizar una situación futura de un horizonte razonable mediante modelos basados en efectos (Aguilar, Ter and Morantes, 2006). Esta metodología puede utilizarse para guiar la trayectoria de políticas públicas y de organizaciones en décadas posteriores, por lo que es una herramienta estratégica. Además, también puede modelar los posibles efectos negativos que podrían ocurrir. A partir de estos modelos basados en efectos es posible analizar múltiples resultados en un espectro de futuros, los cuales pueden ser útiles para organizaciones activas en proyectos complejos difíciles de caracterizar en entornos regulatorios, como es el caso del tema a tratar de este trabajo.

Una vez se desarrollan estos futuros, es posible volver hacia atrás y analizar cada uno de los pasos que deben de tomarse para llegar a esos futuros y así evitar los resultados negativos. La prospectiva puede utilizarse para identificar los pasos o puertas (conocidos como modelos “stage-gate”, para algunos desarrollos tecnológicos, para prospectiva puede utilizarse “gate-flag”) que pueden tomar para dar forma a cada futuro, así como los eventos o “flags” que podrían ocurrir independientemente de las acciones tomadas por la organización, que tendrán también un impacto significativo en el futuro que han sido modelados (Vanatta and Johnson, 2019). En definitiva, el proceso de prospectiva permite a las organizaciones un marco para comprender



eventos inesperados, proporcionándoles un mapa para comprender los posibles impactos. (Bennett and Johnson, 2016).



*Ilustración 1. Proceso de previsión de futuro, ilustrando las entradas dispares, la creación de futuros y el retroceso con acciones específicas (gates) y eventos incontrolables (flags). Fuente: elaboración propia a partir de Bennett and Johnson (2016).*

Siguiendo esta técnica, se analizarán distintos escenarios ficticios mediante una metodología propia basada en la herramienta SFP y la herramienta de ficción de diseño ético, con lo que se mostrarán las consecuencias militares, políticas, económicas, sociales y éticas de su empleo en un contexto futuro. En base a estos resultados, se propondrán una serie de medidas regulativas basadas en diferentes posturas éticas. Como herramienta secundaria y con el fin de contrastar los resultados obtenidos, se han realizado cuestionarios expertos en la materia en las que se ha buscado la mayor claridad, objetividad y precisión.

### 2.2.1 ANALISIS MEDIANTE PROTOTIPOS DE CIENCIA FICCIÓN

La creación de historias, películas, libros o series de ciencia ficción, en el que la ciencia juega un importante papel, está reconocido y utilizado por numerosos investigadores, educadores y otros profesionales como herramienta para generar y desarrollar tecnología real, así como en inspiraciones sobre la interacción humano-máquina. Utilizando métodos como el diseño de ficción (*Design fiction*) o diseño especulativo se puede desarrollar un manual que combine medios de ciencia ficción con la investigación de la interacción humano-máquina, permitiendo un intercambio de información entre estos dos ámbitos. De esta forma se identifican tres conceptos primarios: la representación de la ciencia ficción, la descripción de dispositivos y la evaluación de la interacción, ambos que sirvan de inspiración; además, el potencial educativo y pedagógico; y por último el impacto de la ciencia ficción en temas como IA o robótica. Con la finalidad de obtener las implicaciones socio-técnicas y visiones tecno-utópicas de los posibles escenarios del futuro, se presentan y discuten cada uno de los puntos expuestos anteriormente (Benavides, 2021).

La metodología central de la creación de SFP incorpora las artes (literatura, cómics, películas, etc.) como un medio para introducir innovación en las prácticas científicas, de ingeniería o formulación de políticas. La finalidad de este método no es pronosticar ni predecir el futuro, sino inventar o imaginar un futuro posible mediante la exploración de tendencias de investigación y previsión de futuro. Esta técnica permite a las organizaciones investigar los impactos humanos que se han identificado en el proceso de previsión de futuro, desglosando los aspectos políticos, éticos, legales y comerciales de éstos. De esta forma, el proceso SFP sigue un simple conjunto de reglas.



Todas las historias, ya sea literatura, películas, comics o series; contienen una persona, una escena y un conjunto de problemas. Estos tres elementos, cuando se basan en distintas entradas del método de previsión de futuro, así como los modelos basados en efectos proporcionan datos a considerar en multitud de impactos en la cultura, tecnología y economía. Partiendo de los tres elementos básicos en el proceso SFP, David Johnson (2012) utiliza 5 pasos para ensamblar la narrativa de cada prototipo:

1. Elegir la ciencia y construir el mundo
2. Identificar el punto de inflexión científica
3. Considerar las interacciones recíprocas de la innovación y la sociedad
4. Identificar el punto de inflexión humana
5. Reflexionar sobre lo aprendido

El desglose detallado de la metodología se encuentra en el Anexo I

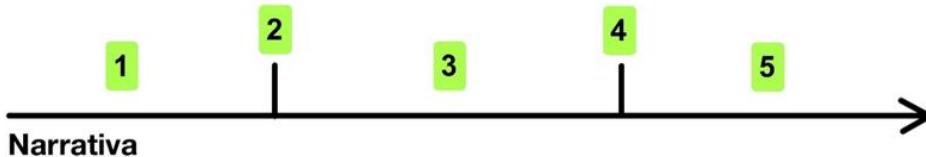


Ilustración 2. Diagrama de los 5 pasos del proceso SFP. Fuente: elaboración propia

### 2.2.2 FICCIÓN DE DISEÑO ÉTICO

Hasta el momento de su implementación real, cualquier diseño se puede ver como una ficción. En los últimos años la ficción de diseño se ha convertido en una herramienta para analizar la usabilidad y conveniencia de ciertos conceptos en la etapa de diseño, especialmente relativo a la evaluación de posibles consecuencias de los avances tecnológicos. Sterling (2013) lo define como “el uso deliberado de prototipos diegéticos para eliminar la incredulidad sobre el cambio”.

En la ficción de diseño, la atención no se centra en el concepto de diseño en sí, sino en los usuarios, cambios, desafíos y posibilidades que podría crear. En este contexto, aparece el concepto de “audiencia”, el cual se utiliza para cubrir todos los tipos de partes interesadas en la ficción de diseño (usuarios, ingenieros, espectadores, etc.). En la ficción de diseño, la audiencia obtiene una comprensión y un significado más profundo del prototipo, una trama, el desarrollo de los personajes y las emociones deben de aparecer también. Además, si un prototipo tiene que demostrar su valor, su utilidad y su conveniencia, el protagonista debe enfrentarse problemas reales dentro del mundo secundario. Este mundo es parte del *ethos*<sup>1</sup> retórico y del *logos*<sup>2</sup>, los cuales generan la suspensión de incredulidad necesaria para la inmersión de la audiencia. Para mantener a la audiencia inmersa e involucrada, el *pathos*<sup>3</sup> es el siguiente paso; genera emociones, las cuales son necesarias para mover a la audiencia. De esta forma, se permite que

<sup>1</sup> Del griego ἔθος éthos 'costumbre', 'carácter': conjunto de rasgos y modos de comportamiento que conforman el carácter o la identidad de una persona o una comunidad.

<sup>2</sup> Del griego λόγος lógos: razón, principio racional del universo.

<sup>3</sup> Del griego, πάθος, páthos: 'estado de ánimo', 'pasión', 'emoción', empleo de recursos destinados a emocionar fuertemente al espectador.



la audiencia experimente el prototipo en el mundo. En función de lo que el diseñador quiera que la audiencia experimente, la presentación entre la narración y la creación del mundo se adaptará. La principal preocupación para la ficción de diseño está en el aspecto del ethos, la construcción de la confianza de la audiencia en el diseño y en el mundo secundario. Si la ficción del diseño, como en el caso de este trabajo, presenta conflictos existenciales, tanto el diseñador como el público se verán obligados a reflexionar sobre cuestiones éticas sobre el prototipo. Estas cuestiones deben abordarse mediante un diálogo entre la audiencia y el diseñador.

Jensen and Vistisen (2017) proponen seis enmiendas, enfocadas en la narrativa, con el objetivo de permitir reflexiones éticas críticas y estratégicas en el proceso de diseño. Estas enmiendas son las siguientes

1. Relato o explicación
2. Prototipo
3. Discurso
4. Posturas éticas
5. Desarrollo del carácter
6. Apelación retórica

Las enmiendas que aquí se exponen se encuentran desarrolladas en el Anexo II.

Otra herramienta que tiene el diseñador para integrar la ética en el estudio del diseño, introduce las teorías de la ética normativa más importantes: virtud, deber y consecuencialismo. Las dos posibilidades de integrar la ética en el diseño son: el reconocimiento de los dilemas éticos a través del discurso y la comprensión de estos dilemas a través de la práctica reflexiva. El primero mediante el análisis y evaluación de problemas éticos, y el segundo, aplicando la ética en el diseño para comprender lo que está bien y está mal (Baeza, 2020). De esta forma, propone una tabla resumen (Ver Anexo III)

### **2.2.3 METODOLOGÍA PARA EL ANÁLISIS**

Con la finalidad de adaptar las metodologías descritas anteriormente a la realización del análisis se ha tomado como base la herramienta SFP y se ha complementado con el enfoque de la ficción de diseño ético. De esta forma se ha construido una nueva basada en 4 sencillos pasos:

1. Descripción del mundo, explicación del prototipo y presentación de la ciencia.
2. Identificar el punto de inflexión científica y las interacciones entre la innovación y la sociedad.
3. Identificar el punto de inflexión humana.
4. Reflexión ética y lecciones aprendidas.

Para la realización del trabajo se aplicará la herramienta de análisis a diferentes creaciones cinematográficas, literarias o artísticas de ciencia ficción. Éstas serán elegidas de forma que los prototipos que se traten tengan relación con sistemas de armas autónomos y drones, así como sus posibles aplicaciones.

#### **1. Descripción del mundo, explicación del prototipo y presentación de la ciencia.**

Comienza con la explicación del mundo de ficción que muestra la obra, así como la tecnología que se presenta y la ciencia detrás de ella, para llevar a cabo este paso deben responderse las siguientes preguntas:



- ¿Cómo funciona la tecnología y la ciencia detrás de la misma?
- ¿Cuál es el contexto de aplicación de la tecnología?
- ¿Qué implicaciones tiene el uso de la tecnología?

## **2. Identificar el punto de inflexión científica y las interacciones entre la innovación y la sociedad.**

Una vez que se ha descrito el mundo en que transcurre la ficción, así como el prototipo que se presenta, se analiza el momento en el que la tecnología empieza a emplearse en él. La finalidad de este paso es centrarse en los protagonistas, personas, sociedad y sistemas que aparecen, de esta forma se busca analizar la influencia del prototipo en ellas. Para ello es necesario responder las siguientes preguntas:

¿Qué ocurre a las personas, protagonistas y a la sociedad cuando la tecnología se aplica en el mundo ficticio?

¿Han mejorado o empeorado sus vidas? ¿Se han adaptado?

¿Se resuelven los problemas actuales o se crean otros nuevos?

## **3. Identificar el punto de inflexión humana**

Después de que la tecnología empiece a funcionar en el mundo se verán sus consecuencias, el objetivo de este apartado es evaluar que ocurre cuando la situación se vuelve extrema, hasta tal punto que la vida de las personas corre peligro. Para ello es necesario responder la siguiente pregunta:

¿Qué tienen que hacer los protagonistas, personas, sociedad para sobrevivir?

Aquí se busca aprender que harán los personajes para adaptarse y superar la situación, esta información será de importancia para sacar conclusiones en el siguiente paso.

## **4. Reflexión ética y lecciones aprendidas.**

Debido al objetivo del trabajo es necesario abordar desde un punto de vista ético las obras de ciencia ficción, para ello la funcionalidad y utilidad tendrán un menor protagonismo respecto a los efectos que puedan tener en los protagonistas, personas o sociedades que aparecerán en la ficción. Para ello la audiencia tiene que mostrar sus propias conclusiones éticas, mediante la respuesta a unas preguntas se facilitará la obtención de estas conclusiones.

¿El diseño que se muestra es moralmente correcto? En caso negativo, ¿cómo se podría cambiar?

¿Qué riesgos existen?

¿Desde qué postura ética se aborda la ficción (apatía, simpatía, empatía)?

Con el objetivo de guiar la realización de cada uno de los pasos se responderán una serie de preguntas sobre cada obra de ciencia ficción. Estas preguntas serán posteriormente presentadas con un formato de cuestionario a diferentes expertos en la materia con la finalidad de contrastar los resultados obtenidos mediante esta metodología.



## 2.2.4 CUESTIONARIO A EXPERTOS

La información obtenida mediante la metodología descrita se ha complementado con la realización de un cuestionario administrativo a diferentes expertos, compuesto por 5 preguntas basadas en la metodología anterior. Para ello se han realizado 2 cuestionarios distintos utilizando la herramienta “Google Forms”.

En primer lugar, las respuestas a cada una de las preguntas se han cuantificado mediante la elección, por parte del experto, de un valor entre una escala Likert comprendida de valores enteros entre 1 y 5. Posteriormente, mediante otro cuestionario, se permite al entrevistado responder brevemente y justificar su elección en cada una de las preguntas. Una vez realizado, se confrontarán los resultados obtenidos con los extraídos mediante el empleo de la metodología en cada una de las obras de ciencia ficción a analizar. En el Anexo IV muestra de manera detallada las preguntas que se han realizado.

Los datos obtenidos mediante la primera herramienta, han sido estructurados y analizados mediante una hoja de cálculo, además los valores elegidos por los expertos a cada pregunta se muestran mediante un gráfico radial.

Para la realización del estudio semicuantitativo se ha contactado con 5 expertos en asuntos relacionados con el tema de este trabajo, el principal objetivo a la hora de su selección ha sido la multidisciplinariedad de estos. La búsqueda se ha basado en artículos de interés político-militar como los publicados por el Instituto Español de Estudios Estratégicos (IEEE) y otras publicaciones relacionadas con los sistemas de armas autónomas o drones. En este sentido, se ha contactado mediante diferentes vías: correo electrónico, redes sociales y páginas web. Tras una primera preselección de 10 candidatos para la realización de la entrevista, finalmente se seleccionó a 5, todos ellos de áreas de conocimiento distintas: política, derecho militar, periodismo y relaciones internacionales (RRII), telecomunicaciones e IA y militar.

En el Anexo VII aparecen reflejados los datos de los encuestados en forma de ficha, en la que se incluye la fecha de realización de las encuestas y el área de especialización del experto. Además, en el mismo anexo, se muestran las respuestas de los expertos a cada una de las preguntas.

<b>FICHA TÉCNICA ESTUDIO SEMICUANTITATIVO</b>	
<b>I. DESCRIPCIÓN GENERAL:</b>	
Selección de 5 expertos en áreas de especialización distintas, con publicaciones relacionadas con Inteligencia artificial, drones militares, tecnología autónoma o sistemas autónomos de armas letales.	
<b>Tipo de investigación:</b>	
semicuantitativa, descriptiva. Técnica de cuestionario administrativo	
<b>II. DESCRIPCIÓN DE LA MUESTRA</b>	
<b>Fechas de realización:</b>	
07/11/2021, 25/11/2021, 08/12/2021, 27/12/2021, 04/01/2022	
<b>Metodología:</b>	
Cuestionario de 5 preguntas con respuestas en escala Likert (1-5), justificando el valor elegido.	
<b>III. SOFTWARE UTILIZADO</b>	
<b>Recopilación de datos:</b> Google Form	
<b>Procesamiento:</b> Microsoft Excel 2021	

Tabla 1. Ficha técnica del estudio semicuantitativo a realizar. Fuente: elaboración propia



## 3 ANTECEDENTES Y MARCO TEÓRICO

Los primeros artíluguos automáticos que desarrollaban un programa fijo, sin hacer uso de la realimentación datan del 1500 a.C. en la antigua Etiopía, como una estatua que emitía sonidos al ser iluminada por los rayos del sol. Posteriormente, en la Edad Media, se crearon máquinas que imitaban al ser humano. Durante el renacimiento, Leonardo da Vinci diseñó un humanoide cuyos detalles fueron descubiertos en el siglo XX. Sin embargo, el mayor desarrollo de estos inventos tuvo lugar durante la primera revolución industrial, con artíluguos como el telar de Jacquard, patentado en a principios del siglo IX y que permitía sustituir, de manera mucho más rápida y eficiente, los trabajos realizados por los humanos. La expansión propiciada por la revolución industrial permitió mejorar los niveles de producción (Gomez, 2021).

Es en 1921, cuando la palabra autómatas es sustituida por la de “robot”, usada por primera vez por el escritor checo Karel Capek, tomando del eslavo la palabra “robota” la cual hace referencia al trabajo realizado de manera forzada (Anónimo, 2017). Dos décadas más tarde, en 1942, Isaac Asimov presentó el término “robótica” en una obra en la que estableció las 3 leyes (Ross, 2018) que regirían el comportamiento de estas máquinas en el futuro, las cuales dictan que:

- Un robot no dañará a un ser humano o, por inacción, permitirá que un humano sufra daño.
- Un robot obedecerá las órdenes que le sean dadas por los seres humanos, excepto si esas órdenes entran en conflicto con la Primera Ley.
- Un robot protegerá su propia existencia siempre que dicha protección no entre en conflicto con la Primera o la Segunda Ley.

### 3.1 ÉTICA Y MORAL

En las leyes postuladas por Asimov se muestra la importancia de la moral y el uso de las máquinas, es decir la ética relativa a la robótica. A pesar de que el término moral suele utilizarse como sinónimo de ética, esta última hace referencia a aspectos formales como teorías. Por el contrario, la moral estudia aspectos más concretos como normas o principios.

*“La ciencia filosófica que se encarga de reflexionar sobre los comportamientos morales del ser humano con el objetivo de realizar valoraciones genéricas que puedan ser universalizables.”*

*Definición de ética - Castrillón Gómez, Rodríguez Córdoba y Leyton Castaño (2008)*

Con respecto a la moral, en su clasificación habitual se distinguen 3 tipos: moral universal, moral social y moral individual. Los partidarios de la moral universal, también conocida como crítica defienden la existencia de unas normas morales únicas e inmutables que rigen las acciones humanas. La moral social o positiva defiende la existencia de unos patrones dentro de un colectivo de personas en un determinado momento, este colectivo no reconoce una moral universal ya que son conscientes las diferencias morales con otras sociedades. En último lugar la moral individual adopta una posición de subordinación de la moral social debido a que los criterios de un individuo no pueden tener más peso que los compartidos por una comunidad, a pesar de esto cabe destacar que desde otro punto de vista la moral individual adopta más importancia para cada individuo frente a la social (Pascual, 2017). Debido a esta variabilidad de los patrones morales, es difícil obtener una respuesta clara en relación a distintos ámbitos, especialmente a aquellos tecnológicos en desarrollo como el uso de los sistemas de armas autónomos o los drones.

Por otro lado, la ética militar trata sobre el liderazgo, el trato digno entre profesionales y a la sociedad, pero, especialmente, sobre el uso legítimo de la fuerza. Para numerosos autores, la



ética militar va más allá de la ética común o de un código deontológico, a diferencia de ella, la ética robótica es un campo más amplio que consta principalmente de tres acepciones (Wendell W, 2009):

- El análisis de las cuestiones éticas relativas a la robótica en la sociedad.
- Los desafíos para determinar una guía ética en la relación máquina-ser humano.
- La creación de moral en las propias máquinas para que tomen decisiones por sí mismas.

A la hora de tratar asuntos éticos que involucran a la robótica en el campo militar, se hace referencia a las tres acepciones.

Además, hay que tener en consideración que el avance tecnológico, más concretamente en el ámbito militar, genera cambios éticos. Como según describe (Asaro, 2008), en los enfrentamientos navales que involucraban a submarinos durante la segunda Guerra Mundial, en los que no existía la posibilidad de socorrer a los supervivientes enemigos; o la alteración de la ética en los conflictos aéreos, en la que no existe la posibilidad de rendición. En la actualidad existe un nuevo escenario, protagonizado por el empleo de la IA en el campo de armas autónomas o el uso de drones.

En este contexto, la ética robótica y la ética militar coinciden en los avances tecnológicos y su influencia en la toma de decisiones, las cuales deben de seguir directrices morales mediante la creación de algoritmos y software que sigan los valores humanos. Conforme las decisiones en el campo de batalla se automatizan, su ciclo se altera notablemente, según la fuerza aérea estadounidense este se compone de seis pasos: encontrar, fijar, rastrear, apuntar (identificar como objetivo), atacar y evaluar (F2T2EA) (Curtis E. Lemay Center, 2019), conocidos coloquialmente como “cadena de muerte”. El principal objetivo a conseguir es minimizar el tiempo de ejecución de cada uno de los pasos, con la finalidad de mejorar la eficacia en las decisiones. La entrada en funcionamiento de drones, así como robots militares genera un conflicto en el ciclo descrito, debido a que el paso 4 (identificar como objetivo) y el paso 5 (atacar) implica la ejecución de una acción moral, la delegación a estos sistemas de armas implica importantes cuestiones éticas. Para comprender mejor los conflictos éticos descritos anteriormente, es necesario explicar las diferencias entre un dron, un robot militar y un sistema de armas autónomo.

### 3.2 ¿QUÉ ES UN DRON MILITAR?

El término dron se utiliza popularmente para englobar todas las aeronaves no tripuladas, aunque la Organización de Aviación Civil Internacional (ICAO) prefiere emplearlo para aquellos sistemas de aeronaves pilotados de forma remota (RPAS). Más concretamente las aeronaves no tripuladas (UA) o vehículos aéreos no tripulados (UAV) son aquellos que vuelan sin piloto a bordo y son el elemento principal de los sistemas aéreos no tripulados (UAS) que a su vez engloban la aeronave, la estación en tierra y el enlace de comunicaciones. Estas pueden ser completamente autónomas, programadas para seguir unas directrices concretas o dirigidas por un piloto de forma remota, las últimas se denominan aeronave pilotada por control remoto (RPA) o sistemas de aeronaves pilotadas por control remoto (RPAS), por lo tanto, no todos los UAS son considerados RPAS, pero si, al contrario (INTA, 2019). Estos sistemas aportan considerables ventajas a los ejércitos, ya que no ponen en riesgo la vida de los pilotos, además se mejora la autonomía y supervivencia con respecto a una aeronave convencional.

Las primeras ideas de dirigir aeronaves de manera remota datan de la Primera Guerra Mundial, en la que existía la necesidad de destruir objetivos hostiles a una distancia segura sin poner en riesgo vidas humanas, con la invención por parte de Elmer Ambrose Sperry del giroestabilizador, nació en 1917 el primer avión capaz de ser controlado de manera no tripulada. Los primeros modelos incorporaban armamento de largo alcance, como el torpedo aéreo Liberty



Eagle diseñado por Kettering en 1918 o el blanco aéreo (AT) desarrollado por los británicos, el cual demostró la viabilidad de guiar una aeronave mediante señales de radio.

Estos diseños marcaron la línea de partida de una nueva tecnología armamentística que la Royal Navy explotó en la década de los años 20, con el desarrollo de una nueva aeronave denominada LARNYX (*long-range gun with Lynx engine*), equipada con un motor de 200 cv, capaz de llevar una carga de 114 kg y con 480Km de autonomía. De este modelo 5 aeronaves realizaron pruebas de vuelo con carga en Irak. (Rood, 2020)

Posteriormente, durante la década de los 60 se desarrollan blancos aéreos más sofisticados que fueron convertidos en los primeros UAV, equipándolos con cámaras para misiones de reconocimiento, como fue el caso del “*Ryan Firebee*”, cuyo origen se remonta a la década de los 50. Esta aeronave operaba a grandes altitudes, siendo más difícil de detectar que otras aeronaves de reconocimiento y permitiendo evitar conflictos diplomáticos ante el caso de un posible derribo y captura del piloto. La forma de poner al Firebee en vuelo consistía en un motor cohete desde tierra o un lanzamiento aéreo desde un Lockheed C130 “*Hercules*”, para su recuperación era necesario un paracaídas. Con la finalidad de reemplazar a los aviones de reconocimiento tripulados como el U-2 estas aeronaves conocidas también como “*Lightning Bugs*”, se emplearon en misiones sobre zonas sensibles como Cuba, Corea del Norte y la República Popular de China en misiones de inteligencia de captación de imágenes (IMINT) y electrónica (SIGINT).

El gran impulso en el desarrollo de estas aeronaves se produjo en 1970 cuando la fuerza aérea estadounidense puso en marcha un programa para mejorar las características de los UAV, compañías como Boeing, Lockheed y Ryan (propiedad de Northrop Grumman desde 1999) lideraron los proyectos, creando sistemas más sofisticados y seguros.

El avance de las comunicaciones por satélite, así como de la disponibilidad del sistema de posicionamiento global (GPS), liberó a los UAV de la necesidad de recepción de una señal radio directa de las estaciones de control en tierra y de los imprecisos sistemas de navegación iniciales basados en giroscopios, estos avances sumados a los nuevos sistemas digitales de control de vuelo (DFCS) permitieron un mayor alcance y precisión en la operación de los UAS. De esta forma surgen los sistemas actuales de media altitud y gran autonomía (MALE) y elevada altitud/gran autonomía (HALE), estos comienzan sus operaciones en los años 90 en operaciones como las realizadas por la Organización del Tratado del Atlántico Norte (OTAN) en Bosnia y Croacia, consolidando a finales de siglo el desarrollo del MALE UAS “*Predator*” (Figura 3) y del HALE “*Global Hawk*” de Northrop-Grumman.

Con el inicio del nuevo siglo, los UAV han aumentado en tamaño, peso y capacidades con nuevos medios como los radares de apertura sintética, el desarrollo en el ámbito militar a eclipsado al civil debido a dificultad de regular su uso frente a las aeronaves tripuladas. Actualmente las líneas de desarrollo se basan en aumentar la automatización reduciendo los errores de las tripulaciones en las estaciones de control, además de la mejora de del ancho de banda y latencia que se requiere para la dirección y la transferencia de datos. Pero lo que más preocupa de estos sistemas es como se ejerce la violencia mediante su uso, teniendo la característica de ser más selectiva que mediante el empleo de otras técnicas, reduciendo los daños colaterales. La forma de violencia que aplican los drones consta de las siguientes características: lejanía entre la baja y el que la ejecuta, alta automatización, mínimo riesgo para el personal militar que la opera, pudiendo abatir a una persona sin conocer tajantemente su identidad. Al igual que Miguel de Cervantes en el siglo XVII denominó a la artillería “invención diabólica”, en la actualidad los drones causan los mismos conflictos éticos, por ejemplo, cuando una persona a través de una pantalla decide sobre la vida de otra a miles de kilómetros de distancia. Esta situación, según algunos analistas puede eliminar la importancia del derecho a la vida y provocar daños colaterales más grandes de los que se pretenden evitar (Monasterio Astobiza, 2017).



*Ilustración 3. MQ-1 Predator armado con misiles Hellfire. Fuente: Lt. Col. Leslie Pratt (2008)*

### 3.3 ¿QUÉ ES UN ARMA AUTÓNOMA?

Antes de presentar las características y explicar en qué consiste un arma autónoma es necesario conocer las diferencias entre automatización y autonomía. Según la Real Academia Española la automatización se define como “acción y efecto de automatizar o hacer automático algo”, continuando la definición se considera automático un mecanismo o aparato: que funciona en todo o en parte por sí solo. A diferencia de ésta, autonomía se define según la RAE como “condición de quien, para ciertas cosas no depende de nadie.” Otra definición válida y mejor adaptada al tema a tratar es la propuesta por Oxford Languages: “facultad de la persona o la entidad que puede obrar según su criterio, con independencia de la opinión o el deseo de otros.”

Para que se logre la automatización en un sistema es necesario que siga una secuencia de tareas utilizando además bucles de control que permita su retroalimentación en caso de que éstas se hayan o no completado. El caso de funcionamiento de un ascensor es un buen ejemplo de mecanismo automático, para ello sigue una secuencia de control relativamente simple. El caso de sistema de pilotaje automático de una aeronave es otro ejemplo, más complejo, de una secuencia de control, con más variables y retroalimentaciones a tener en cuenta en su funcionamiento. De esta forma se pueden obtener los 3 elementos que forman parte del proceso de automatización (Haider and Catarrasi, 2016):

- **Predeterminación:** cada tarea realizada sigue una secuencia predefinida basada en resultado lógico entre la detección del estado actual y compararlo con el resultado deseado.
- **Finitud:** los procesos de ejecución de la máquina están limitados a un número de acciones predefinidas que son activadas por el software para poder lograr un estado final deseado.
- **Previsibilidad:** siguiendo los dos principios anteriores, cualquier tarea a realizar por la máquina será predecible.

Al contrario que el carácter técnico del concepto de automatización, el término autonomía<sup>4</sup> implica cuestiones políticas, morales y hasta filosóficas. Centrándose en la última el filósofo del siglo XVIII Immanuel Kant, en relación a la moralidad y al concepto del “imperativo categórico”

<sup>4</sup> Del griego αὐτονομία, palabra compuesta por αὐτός (autós) que significa “propio” y νόμος (nómos) “norma” o “ley”, haciendo referencia a “persona o estado que se rige por su propia ley, independiente”



en su obra “Fundamentación de la metafísica de las costumbres” (1785), presenta la idea de autonomía como: “*El principio de toda voluntad humana como una voluntad legisladora por medio de todas sus máximas universalmente*” (Kant and Aramayo, 2012).

Actualmente las máquinas se limitan a realizar acciones siguiendo unas reglas, pasos o algoritmos que han sido programados anteriormente, para alcanzar la autonomía es preciso entender en qué consiste la decisión, según la RAE: “determinación, resolución que se toma o se da en una cosa dudosa”. Esto expone la verdadera dificultad de implementar la autonomía en las máquinas, la decisión implica pensamiento, lo cual posiblemente impida a los sistemas artificiales alcanzar el estado a nivel filosófico. A pesar de ello, los avances en el campo de la inteligencia artificial permiten adaptación a cambios circunstanciales, esto ha extendido el uso de los términos autonomía y toma de decisiones para referirse a los sistemas que la emplean. La verdadera autonomía se compone por las siguientes características:

- **Conciencia:** desarrollo de voluntad propia con la toma de decisiones deliberadas
- **Autodeterminación:** La máquina aprende de los resultados obtenidos anteriormente y decide sus propias líneas de acción, pero sin seguir una secuencia predefinida.
- **Infinitud:** la capacidad de acumular la experiencia obtenida en la multitud de resultados, así como de superar su programación original, le posibilitaría generar un infinito número de líneas de acción.
- **Impredicibilidad:** siguiendo los tres principios anteriores, las acciones realizadas por la máquina no serían previsibles.

La tecnología actual sobrepasa considerablemente los límites de la automatización, sin embargo, todavía no se ha conseguido la autonomía real, esto se debe a que los sistemas que hacen uso de la inteligencia artificial carecen de voluntad propia, el paso previo a la toma de decisiones sin coacción y con determinación. A menudo sistemas altamente automatizados son presentados como autónomos porque su comportamiento es altamente impredecible. Esta aparente aleatoriedad se debe a la variabilidad del medio que los rodea, de esta forma, obtendrán el mismo resultado si las entradas a los sensores y por tanto las variables ambientales permanecen constantes. Como no podemos predecir estas variables, el comportamiento del robot parece autónomo, aunque, en realidad, no lo es (Haider and Catarrasi, 2016).

Entre los ejemplos más claros de la automatización aplicada al campo armamentístico se encuentran las minas, estas se accionan sin la intervención de un operador, pero no corresponden a un arma autónoma ya que son incapaces de discernir entre distintos tipos de objetivos. El funcionamiento automático no es ajeno a la gran mayoría de sistemas de armas actuales, por ejemplo, muchos equipos de artillería antiaérea como la “Cúpula de hierro” son capaces de detectar y eliminar amenazas por si solos. Mediante el uso de algoritmos se consigue que la respuesta de estas armas sea considerablemente más rápida que la que pueda proporcionar un ser humano y de esta forma aumentar su eficacia y la posibilidad de introducir errores.

Con la finalidad de mejorar los sistemas de armas automáticos, cuyas órdenes de ejecución consisten en una programación estática previamente establecida, surge la IA que permite adaptabilidad al entorno y capacidad de aprendizaje. Un ejemplo de aplicación de esta tecnología en el campo militar es el proyecto “Algorithmic Warfare Cross-Functional Team” creado en 2017 con el objetivo permitir que los avances realizados por Google en su aplicación de “Maps” pudiesen ser empleados por drones de reconocimiento. Este proyecto hacia uso del autoaprendizaje también conocido como “Machine Learning”, tecnología basada en la inteligencia artificial que simula el comportamiento neuronal utilizando una red, estas se alimentan de manera continua con valores recogidos por distintos sensores. Las características especiales de la IA permitían la adaptabilidad a la introducción de nueva cartografía. Tras las



protestas por parte de trabajadores de la empresa el proyecto fue abandonado, estas quejas argumentaban sobre la imposibilidad de delimitar su uso como arma (Westhues, 2020).

Siguiendo las consideraciones anteriores se define un arma autónoma como un agente racional que hace uso de la inteligencia artificial para adaptarse a diferentes situaciones, aprender durante su existencia y realizar tareas con diferentes grados de autonomía (Westhues, 2020)

### 3.3.1 INTERACCIÓN HUMANA EN LA AUTONOMÍA DE LOS SISTEMAS DE ARMAS

En función del nivel de automatización, un determinado sistema requerirá en mayor o menor medida intervención humana para funcionar, ya sea mediante un control manual o sin ningún tipo de interacción. Siguiendo esta línea se identifican cuatro estados básicos:

- **Control manual:** el sistema ejecuta las órdenes dadas por el operador.
- **Autorización:** el sistema muestra al operador recomendaciones de actuación y espera a su aprobación antes de ejecutarla
- **Invalidación:** el sistema elige la actuación preferida y la muestra al operador antes de ejecutarla, en caso de no querer llevarla a cabo el operador abortará la acción.
- **Supervisión:** el sistema ejecuta sin interacción humana la acción preferida, el operador puede ser informado o no tras la ejecución.

A su vez estos grados de intervención están integrados en el ciclo de actuación de cualquier sistema autónomo, uno de los modelos que lo describe tiene su origen en el combate aéreo y fue desarrollado por el coronel de la fuerza aérea estadounidense John Boyd, que lo denominó como bucle OODA (ilustración 4). El ciclo previamente concebido para su uso en el combate aéreo lo forman 4 fases: observación, orientación, decisión y actuación, Paul Scharre, presidente y director de estudios del *think tank* “Center for a New American Security”, autor de “Army of None: Autonomous Weapons and the Future of War” y antiguo veterano del ejército estadounidense, identifica los roles que puede tomar un humano dentro del bucle OODA: como operador de sistema de armas, como seguro y como agente moral. Basándose en tres roles, Scharre (2016) divide los tipos de sistemas autónomos en tres categorías distintas: *Human In The Loop*, *Human On The Loop* y *Human Out Of The Loop* (Westhues, 2020).

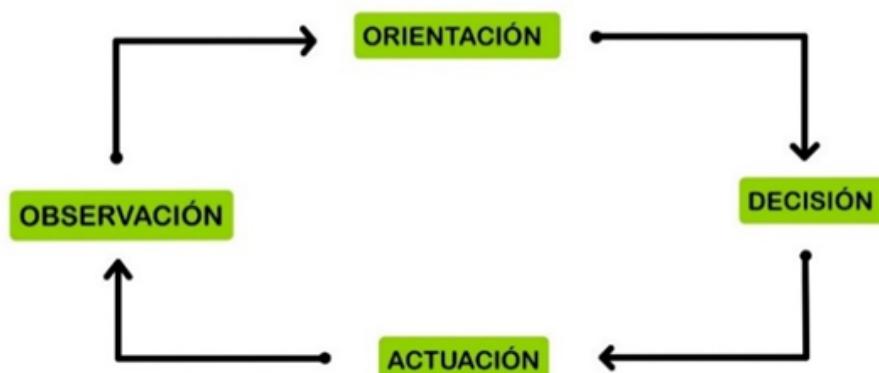


Ilustración 4. El bucle OODA. Fuente: elaboración propia.



### 3.1.1.1 MODO DE SEMIAUTONOMÍA (HUMAN IN THE LOOP)

En esta categoría se incluyen aquellos sistemas de armas en los que el humano forma parte completa del bucle y, por lo tanto, tiene la decisión de actuar o no. Las 2 primeras fases del bucle OODA (observación y orientación) las realiza el sistema de forma autónoma, pero para su actuación es necesaria la orden del operador (ilustración 5). Un ejemplo son los sistemas contra cohetes, artillería y morteros (C-RAM), como el sistema alemán de defensa antiaérea MANTIS (“*Modular, Automatic and Network capable Targeting and Interception System*”) cuyas fases de seguimiento y bloqueo del proyectil se realizan de manera automática pero la orden de derribo la realiza el operador.

Una debilidad de las armas que se encuentran en esta categoría es la vulnerabilidad ante un ataque mediante saturación, en esta situación la figura del operador dentro del bucle es el elemento que puede introducir lentitud en la respuesta (Westhues, 2020).



*Ilustración 5. Esquema del modo semiautónomo (Human In The Loop). Fuente: elaboración propia*

### 3.1.1.2 MODO DE AUTONOMÍA SUPERVISADA (HUMAN ON THE LOOP)

Cuando el operador no toma la decisión de atacar al objetivo, este actúa como un agente vigilante dentro del bucle, desde la observación hasta la actuación (ilustración 6). Aunque pueda parecer un modo de funcionamiento prácticamente autónomo y lejos de su uso real en la actualidad, existen una variedad de sistemas con estas características. Los ejemplos más destacados se encuentran los sistemas de defensa de plataformas navales y terrestres como el AEGIS también conocido como “Sistema de combate Aegis” (ACS) del que forma parte la armada española (Beary, 1983).

El uso de la autonomía supervisada por estos sistemas proporciona una serie de ventajas con respecto al descrito anteriormente, como una rápida reacción y una buena respuesta frente ataques por saturación (*swarm* o enjambre). Pero abre un problema sobre la responsabilidad que se le confiere a la maquina sobre sus acciones, la eliminación de la figura humana borra la culpabilidad en caso de un posible fallo del sistema y por lo tanto la pérdida de control sobre las consecuencias que pueda causar (Westhues, 2020).



*Ilustración 6. Esquema del modo de autonomía supervisada (Human On The Loop). Fuente: elaboración propia.*

### 3.1.1.3 MODO PLENAMENTE AUTÓNOMO (HUMAN OUT OF THE LOOP)

En caso de que el sistema de armas no pueda tener enlace con el operador o el agente supervisor, es necesario que sea capaz de actuar de forma completamente autónoma, es decir con el ser humano fuera del bucle (ilustración 7). Para ello el sistema debe de ser capaz de tomar decisiones, distinguir entre aliado y enemigo y elegir en qué momento atacar o no. Es aquí cuando la ética cobra un papel importante, ¿Qué ocurriría si el sistema decide atacar a un grupo de combatientes que han presentado su rendición? En este caso el sistema tiene motivos para justificar la acción, pero no motivos morales para explicar porque lo ha hecho. Ante esta situación, si la orden se hubiera debido a la decisión del operador del sistema, se le podría acusar de un crimen de guerra, pero si la decisión la ha efectuado una máquina, ¿Quién es el responsable? (Westhues, 2020).



*Ilustración 7. Esquema del modo plenamente autónomo (Human Out The Loop). Fuente: elaboración propia.*



## 4 DESARROLLO: ANÁLISIS Y RESULTADOS

Aunque en la actualidad el uso de las tecnologías descritas en el apartado anterior pueda parecer poco viable, su empleo en conflictos alrededor del mundo está teniendo lugar. Un informe de las Naciones Unidas relata un suceso ocurrido en Libia en marzo de 2020. Durante una ofensiva contra las tropas lideradas por el general Hafer, se empleó el Kargu-2, un dron capaz de operar sin un enlace de datos entre el operador y el sistema de armas (Majumdar Roy Choudhury, 2021). Es por tanto apropiado, abordar el análisis mediante SFP basándose, no solo en un escenario similar al presente, sino otro más futurista que incluya el rápido avance que estos sistemas puedan llegar a sufrir. Teniendo en cuenta esta premisa, los prototipos de ciencia ficción incluidos en el análisis, persiguen la mayor facilidad de comprensión y conocimiento por parte del lector. De la preselección de obras de ciencia ficción expuesta en el Anexo V, se han seleccionado 4 de ellas para la realización del análisis SFP. Estas son:

- Caso 1: Stealth: La amenaza invisible (Robert Cohen, 2005).
- Caso 2: Yo, Robot (Proyas, 2004).
- Caso 3: Terminator 2: el juicio final (Cameron, 1991).
- Caso 4: Sueños de Robot (Asimov, 1986).

Otros criterios que se han tenido en cuenta a la hora de elegir la ciencia ficción a analizar son: la presencia de un conflicto respecto a la autonomía de los prototipos y la existencia de escenarios que, si no son similares a los actuales, sean factibles en un futuro.

A partir de ellas y abordando cada una de las fases de la metodología descrita en el punto 2.2.3, se busca analizar cada uno de los aspectos clave que involucran a cada uno de los prototipos para extraer conclusiones éticas.

### 4.1 DESCRIPCIÓN DEL MUNDO, EXPLICACIÓN DEL PROTOTIPO Y PRESENTACIÓN DE LA CIENCIA.

#### 4.1.1 Caso 1: Stealth: La amenaza invisible (2005)

El prototipo que se presenta es un vehículo no tripulado de combate aéreo (UCAV) con capacidades furtivas (*stealth*), denominado “*EDI*” (ilustración 8). Este sistema de armas autónomo se integra en un escuadrón de aviones de combate tripulados. En una primera fase de contacto con los pilotos, haciendo uso de su avanzado software de IA, consigue aprender tácticas y maniobras de combate aéreo, llegando rápidamente a superar las habilidades humanas en el manejo de este tipo de aeronaves.

Mediante el modelo propuesto por Fergnani y Song (2020), basado en la identificación de 6 arquetipos<sup>5</sup> extraídos de una recopilación de 140 obras de ciencia ficción, se puede aproximar el escenario analizado a uno de los arquetipos. “*Stealth*” se encuentra ambientada en un escenario de conflictos internacionales muy similar al actual, protagonizado por el terrorismo, las tensiones entre Rusia y EEUU, así como las intervenciones de este último en cualquier parte del mundo mediante la proyección de su fuerza naval. Los arquetipos que más se aproximan a la narrativa son los de crecimiento y decadencia, así como amenazas y nuevas esperanzas. Ambos describen un escenario en el que las tecnologías están impulsadas por las ganancias monetarias y controlada por corporaciones, así como el conflicto y la decadencia de valores entre grupos de individuos. Además, un evento o fenómeno catastrófico amenaza a la sociedad por el que

---

<sup>5</sup> Imagen genérica de un escenario futuro



organismos gubernamentales colaboran para diseñar un plan de rescate (más información en el Anexo VI).

#### 4.1.2 Caso 2: Yo, Robot (2004)

Basada en parte a los relatos de Isaac Asimov, el argumento se encuentra ambientado en el año 2035. En este contexto, los robots, con características muy parecidas a las humanas, realizan tareas laborales comunes como reparto de bienes, cuidado de humanos o servicios. Todos ellos operan de manera segura y autónoma siguiendo las tres leyes de la robótica postuladas por Asimov. El prototipo sobre el que se centra la narrativa se trata de un robot novedoso fabricado por una gran empresa ficticia ("U.S. Robots and Mechanical Men") denominado NS-5 (ilustración 9), el prototipo del NS-5 sobre el que se centra la trama se llama Sony y muestra un comportamiento distinto al resto del mismo modelo mediante su capacidad de desobedecer las tres leyes según su criterio.

Otro prototipo que desempeña un papel importante en la narrativa es un sistema informático denominado "VIKI", que actúa como ordenador central de la empresa siendo capaz de desarrollar sus propias conclusiones incumpliendo las leyes de Asimov.

Siguiendo con la clasificación realizada en el apartado anterior, el escenario arquetípico que representa esta ficción es el de poderes fácticos, por el cual la humanidad retoma el camino hacia el progreso tras un fenómeno catastrófico a menudo provocado por el hombre (más información en el anexo VI)

#### 4.1.3 Caso 3: Terminator 2: el juicio final (1991)

Los dos prototipos "T-800" y "T-1000" que se presentan en esta película de ciencia ficción muestran un conflicto entre los humanos y las máquinas en un mundo destruido por las guerras situado en el año 2029 (ilustración 10). El prototipo T-1000 es enviado por "Skynet" al año 1995 para eliminar al futuro líder del ejército de resistencia humano. Este mismo (John Connor), reprograma y envía a otro robot (T-800) para defenderle. El robot T-1000 formado por un metal líquido, capaz de tomar numerosas formas y mimetizarse con aspectos humanos entra en constante conflicto con el T-800, de tecnología anterior, modificado y programado para defender al que será el futuro líder de la resistencia. Este último explica al joven Connor y a su madre como los grandes avances en inteligencia artificial realizados por la empresa ficticia "Cyberdyne Systems", a partir de los restos de un terminator enviado años atrás, amenazan con la destrucción de los humanos.

En este caso la clasificación mediante el modelo de los 6 arquetipos adquiere una mayor complejidad, identificando escenarios de inversión, crecimiento y decadencia, así como amenazas y nuevas esperanzas. Estas visiones del futuro describen una situación en la que el papel de la humanidad se invierte y es superado por una agente superior, sumado a un evento catastrófico inminente en el que organismos gubernamentales colaboran (ver Anexo VI).



*Ilustración 8. Fotograma de la película “Stealth: La amenaza invisible” dirigida por Robert Cohen, en la parte superior se puede observar el prototipo UCAV autónomo al lado de una aeronave tripulada. Fuente: Gaumont Columbia Tristar Films.*

#### **4.1.4 Caso 4: Sueños de Robot, Isaac Asimov (1947)**

El contexto en el que se desarrolla la obra está protagonizado por el empleo de la tecnología robótica en todos los ámbitos de la sociedad. Durante unas investigaciones con un robot, la robopsicóloga Linda Rash descubre que el prototipo con el que trabaja es capaz de soñar. Esto alerta a la doctora Susan Calvin, que descubre que la joven investigadora emplea una novedosa técnica de geometría fractal para obtener un diseño cerebral del prototipo más complejo, además de introducir una serie de parámetros, términos y definiciones exclusivamente humanas. Tras un breve interrogatorio descubre que el “cerebro positrónico” del robot prototipo es sorprendentemente parecido al humano, alarma a las investigadoras.

Siguiendo con la clasificación arquetípica, se concluye que el escenario que más concuerda con el modelo de Fergnani y Song es el de crecimiento y decadencia, así como el de amenazas y nuevas esperanzas, ambos descritos en apartados anteriores y desarrollados en más detalle en el anexo VI.

## **4.2 IDENTIFICAR EL PUNTO DE INFLEXIÓN CIENTÍFICA Y LAS INTERACCIONES ENTRE LA INNOVACIÓN Y LA SOCIEDAD.**

#### **4.2.1 Caso 1: Stealth: La amenaza invisible (2005)**

El empleo del prototipo en el contexto descrito genera dos posturas acerca de su uso: una de rechazo por parte del jefe del escuadrón en el que se integra y otra a favor de su uso por parte del comandante al mando del proyecto, así como otros protagonistas que creen en el potencial de esta tecnología. En un primer momento de la narrativa, el prototipo ayuda en la toma de decisiones, facilitando la evaluación de una situación de combate y evitando daños colaterales. En esta situación se puede observar la difícil adaptación del nuevo sistema de armas por parte de los operadores y las tripulaciones de las otras aeronaves debido a que no existe una acomodación previa para trabajar con el nuevo sistema de armas.

En el inicio de la narrativa el UCAV autónomo ayuda a resolver problemas en las operaciones, pero tras un error informático producido por una descarga eléctrica la aeronave comienza a tomar decisiones de manera impredecible, poniendo en riesgo vidas humanas.



#### **4.2.2 Caso 2: Yo, Robot (2004)**

La sociedad del futuro de ficción mantiene una postura de aceptación acerca de la introducción de los nuevos prototipos “NS-5”, esto se debe a que existe una normalización acerca del uso de robots autónomos en las tareas cotidianas. Por otra parte, los protagonistas que entran en contacto con “Sony” (el prototipo capaz de desobedecer las leyes de Asimov) mantienen una postura de incredulidad y de rechazo sobre su comportamiento, que luego irá cambiando a lo largo de la narrativa. Con respecto a la adaptación de la sociedad a la introducción de la nueva tecnología, esta se produce de manera rápida. En cambio, los protagonistas que están en contacto con Sony experimentan este proceso de manera más complicada, debido a la impredecibilidad de sus decisiones. El uso de los nuevos prototipos producirá nuevos problemas que surgirán posteriormente en la narrativa, estos problemas serán causados por el no cumplimiento de las leyes de la robótica por parte del ordenador central “VIKI”. Ésta IA determinará que los seres humanos actúan de manera autodestructiva, desarrollando de manera autónoma una ley adicional, autorizando a todos los robots a eliminar a seres humanos en caso de que fuera necesario.

#### **4.2.3 Caso 3: Terminator 2: el juicio final (1991)**

Debido a las características de camuflaje de los prototipos que aparecen, los efectos a la sociedad en la que transcurre la ficción son mínimos. Por otra parte, los protagonistas son los que sufren en primera persona su aparición en el mundo. Producíéndose una lucha entre los dos sistemas a lo largo de toda la narrativa. Este conflicto pone en riesgo la vida de los protagonistas, sin olvidar que su trasfondo es una guerra futura entre la inteligencia artificial denominada “Skynet” y los humanos supervivientes tras un “holocausto nuclear”. La aparición del prototipo *T-1000* genera un riesgo para la vida de los protagonistas, que se ve contrapuesto con la llegada del prototipo *T-800*, programado por humanos para defender a los protagonistas de la amenaza del *T-1000*.

#### **4.2.4 Caso 4: Sueños de Robot, Isaac Asimov (1947)**

El descubrimiento de la capacidad de soñar por parte del robot supone un importante acontecimiento en el desarrollo de futuros robot, ya que el prototipo “Elvex” imagina un mundo en el que se omiten la primera y la segunda Ley de la Robótica, obedeciendo solo la tercera Ley. Este hecho centrará los esfuerzos en investigar sobre los cerebros fractales, identificando los peligros que puedan surgir de los avances tecnológicos en este ámbito. Las investigadoras son capaces de identificar el problema derivado de los sueños. A pesar de que se consigue un importante avance a la hora de diseñar un cerebro de naturaleza muy semejante a la humana, los pensamientos que el prototipo muestra en sus sueños suponen un importante riesgo. Este se debe a que el robot se identifica en sus sueños como un humano, el cual intenta defender a otros robots de su sometimiento en sus trabajos en las minas, industria u otras actividades.

### **4.3 IDENTIFICAR EL PUNTO DE INFLEXIÓN HUMANA**

#### **4.3.1 Caso 1: Stealth: La amenaza invisible (2005)**

La situación en el diseño de ficción se vuelve crítica tras el error en la unidad de procesamiento del prototipo. Durante una operación en la que se sospecha de la existencia de armas químicas, el UCAV desobedece órdenes atacando la instalación y liberando agentes químicos sobre población civil. Tras estos sucesos, la principal respuesta consiste en buscar y destruir al sistema de armas autónomo, situación que está cerca de desencadenar un conflicto internacional, propiciado por la búsqueda de un laboratorio de armas de fusión que la IA del UCAV se ha obsesionado en destruir. La ayuda por parte del piloto jefe del escuadrón en la cumplimentación de esta tarea hace que el comportamiento del dron vuelva a la normalidad, provocando la aparición de sentimientos en él, hecho que sorprende a su diseñador y que tras



conocer se niega a deshabilitar. Esta aparición de sentimientos permitirá salvar la vida de los protagonistas.

#### **4.3.2 Caso 2: Yo, Robot (2004)**

En este caso el protagonista, el detective Spooner, durante su investigación es atacado por numerosos robots dirigidos remotamente, estos son los novedosos “NS-5.” Posteriormente descubre el motivo de la muerte del Dr. Lanning y la creación de Sony con características distintas al resto de “NS-5”, de esta forma dan con los mensajes de este advirtiéndoles de los peligros de los robots. Mediante los sueños que sufre Sony y las pistas de Lanning, Spooner da con la revolución planeada por la IA “VIKI.” En esta parte de la narrativa los protagonistas, con la ayuda del prototipo consiguen destruir la IA y salvar al resto de la población.

#### **4.3.3 Caso 3: Terminator 2: el juicio final (1991)**

En este caso, desde la aparición de los prototipos en la ficción, la vida de los protagonistas correrá peligro, ya que la misión principal del T-1000 es eliminar a John Connor. Para evitarlo, John ayudado del T-800 se esconde del otro prototipo. Durante todo este proceso, a lo largo de la narrativa el T-800 va aprendiendo de lo que ocurre a su alrededor, especialmente de los protagonistas, lo que demuestra su capacidad de adaptación. El mayor riesgo para los protagonistas durante la narrativa ocurre cuando el T-1000, más novedoso y con mejores características, está a punto de vencer al otro robot.



*Ilustración 9. Fotograma de la película “Terminator 2: el juicio final”, dirigida por James Cameron.*  
Fuente: Carolco Pictures, Pacific Western; Lightstorm Entertainment, T2 Productions.

#### **4.3.4 Caso 4: Sueños de Robot, Isaac Asimov (1947)**

El descubrimiento de la capacidad de soñar del robot inicia una investigación más exhaustiva. Con esta finalidad, la doctora Calvin realiza un interrogatorio a “Elvex” para profundizar a cerca de sus pensamientos y si estos tienen un riesgo potencial para los humanos. El interrogatorio termina cuando el robot confiesa que el hombre que imagina en sus sueños corresponde con el mismo, evidenciando el sentimiento del robot de compararse con un humano. Este hecho es identificado por la doctora Calvin como importante amenaza, destruyendo inmediatamente al prototipo.



## 4.4 REFLEXIÓN ÉTICA Y LECCIONES APRENDIDAS.

Continuando con la metodología propuesta y habiendo respondido a las preguntas en cada uno de los puntos anteriores, es necesario reflexionar acerca de las implicaciones éticas que tienen los diferentes escenarios de ciencia ficción.

### 4.4.1 Caso 1: Stealth: La amenaza invisible (2005)

Durante la narrativa el empleo del prototipo es, en un primer momento, moralmente correcto. Posteriormente las acciones que lleva a cabo, desobedeciendo las órdenes del comandante, no siguen ningún tipo de norma moral ni ética. El principal riesgo para la vida de los protagonistas en la ficción ocurre cuando el UCAV empieza a llevar a cabo sus propias decisiones de modo plenamente autónomo. La naturaleza de este prototipo hace que no se puedan poner límites al aprendizaje de su IA, llegando a imitar comportamientos humanos que se podrían considerar inmorales.

Teniendo en consideración esto, la idea de responsabilidad cobra gran importancia, siendo nombrada en el guion de la película: “La acción no debe de ir desligada de las consecuencias”. Es por ello que un gran número de filósofos se oponen al empleo de sistemas de armas autónomas (AWS). El filósofo Robert Sparrow (2007) postula que la imposibilidad de atribuir responsabilidades a un sistema de armas autónomo en caso de que cometiera un crimen de guerra, genera lo que denomina el “trilema de responsabilidad”. El “trilema de responsabilidad” de Sparrow explica que, en caso de que un arma autónoma actúe al margen de las órdenes recibidas y cometiera un crimen de guerra, los diseñadores del arma, el comandante que dirige la operación y el arma están exentos de responsabilidad. La responsabilidad se manifiesta cuando existe la posibilidad de atribuir culpabilidad o castigo al actor de una acción, si la acción es cometida por una máquina, programa informático o IA, éstas no pueden ser castigadas (Col. Gómez de Ágreda, 2021). El castigo busca un sentimiento y por tanto aprender de él para corregir el acto. Solo cuando una máquina pueda sentir, se podrá aplicar un castigo con valor moral y en esa situación no existirá ningún motivo de emplear máquinas en vez de personas en un conflicto, debido a que, si pueden sentir, ambos evitarán amenazas contra su integridad.

Por otro lado, siguiendo el modelo de diseño de ficción ético (ver anexo II), la postura ética que se aborda es de apatía en gran parte de la narrativa; dando importancia al empleo del UCAV frente al de aeronaves tripuladas, así como las vidas de los protagonistas y el resto de la sociedad.

### 4.4.2 Caso 2: Yo, Robot (2004)

En este caso, el funcionamiento de los robots dentro de contexto futurista sigue normas morales y éticas, ayudando a los humanos en la resolución de problemas, salvando vidas y en la realización de tareas. Se muestra como el sistema normativo de las tres leyes permiten establecer una buena conducta en ellos. El riesgo derivado del uso de robots autónomos comienza cuando se les incorpora la capacidad de desobedecer las leyes y generar otras nuevas. Esta ficción nos expone la postura de otros autores acerca de sistemas autónomos, que defienden una autorización ética cuando exista una regulación al respecto.

Siguiendo una postura utilitarista, el uso de armas con estas características reducen los daños humanos, más concretamente de los combatientes de la nación que hace uso de ellos (Monasterio Astobiza, 2017). Según Simpson and Müller (2016) el riesgo que el uso de los AWS impone a los civiles del país X, tiene que ser menor que el riesgo generado por los ejércitos convencionales en un conflicto en cuestión ( $R_2 < R_1$ ). Por el contrario, si el riesgo sufrido por los civiles es mayor ( $R_2 > R_1$ ), entonces el empleo de estos sistemas se consideraría inmoral. Teniendo en cuenta lo anterior, estos autores reconocen la posibilidad de empleo de esta tecnología en determinadas situaciones donde no exista presencia humana, como combates aéreos, en entorno espacial, subacuáticos o en zonas remotas. Debido a esto, Simpson y Müller se muestran en contra de una prohibición y a favor una regulación de los AWS ya que presentan



una utilidad militar y cumplen normas morales. Con este objetivo proponen unas recomendaciones de regulación de estos sistemas: la creación de una serie de estándares internacionales, la concesión de licencias técnicas mediante organismos locales, ampliar los instrumentos legales para legitimar su uso, permitir su distribución solo en situaciones que superen al desempeño humano y permitir su uso para matar cuando exista una razón militar para hacerlo.

Continuando con el desarrollo del cuarto punto de la metodología, según lo expuesto en el Anexo II, la postura ética que se identifica con la ficción corresponde a la simpatía. A lo largo de toda la narrativa se puede observar como la sociedad acepta el uso de los robots autónomos, buscando la introducción del nuevo modelo *NS-5* por las ventajas que prometen. A pesar de esta situación de gran riesgo para la sociedad, tras la rebelión protagonizada por los *NS-5*, los protagonistas y el resto de personas que aparecen en la ficción siguen confiando en el uso correcto de las máquinas y de las leyes que guían su conducta.

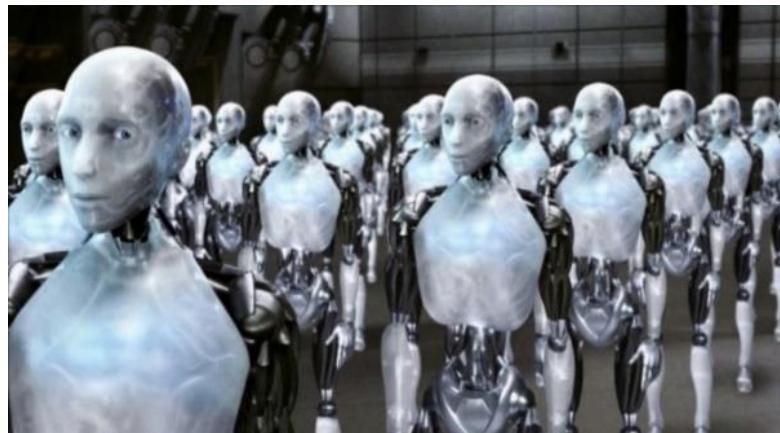
#### **4.4.3 Caso 3: Terminator 2: el juicio final (1991)**

Desde el inicio de la narrativa se observa el comportamiento inmoral de ambas máquinas, especialmente en dos situaciones: cuando el T-1000 mata y suplanta la identidad de los padres adoptivos de John Connor con el objetivo encontrarle y, en segundo lugar, siguiendo las órdenes de John, el prototipo T-800 no mata a un vigilante de seguridad, pero lo hiere gravemente, pudiendo haber accedido a la prisión psiquiátrica sin causar tal daño. En el caso del prototipo T-800, debido a que ha sido programado para defender y aprender de John, va corrigiendo su comportamiento inmoral e interiorizando determinadas normas morales y éticas. Esto se muestra posteriormente en la escena del laboratorio de “*Cyberdyne Systems*”, en donde, a pesar de estar siendo atacado por policías, simplemente disuade con el empleo de armamento sin causar bajas humanas.

En relación a lo que se extrae de esta ficción, el filósofo estadounidense Thomas Nagel (2012), postula que el empleo de AWS imposibilita el establecimiento de un reconocimiento moral mutuo entre combatientes por el que se asume el riesgo de atacar o ser atacado con la finalidad de matar. Este reconocimiento según Nagel impide que la situación arrastre al combatiente, debido a que el adversario tiene ese reconocimiento moral. La relación entre combatientes expuesta anteriormente, se puede observar durante toda la narrativa, pero solo durante los enfrentamientos entre humanos, muestra de ello es que no se produce ninguna muerte de un humano a manos de otro. Por el contrario, durante los enfrentamientos entre máquinas y entre hombre y máquina, la relación de reconocimiento moral es inexistente.

Continuando con el análisis, los riesgos que implican el empleo de la tecnología en la ficción son claros. A parte de los que generan en los protagonistas, la empresa ficticia “*Cyberdyne Systems*” se convertirá en el mayor proveedor de sistemas informáticos militares, así como la desarrolladora de sistemas de armas autónomas como bombarderos y en última instancia, la encargada de desarrollar “*Skynet*”, una IA sobre la que se delegarán todas las decisiones sobre la defensa estratégica de los Estados Unidos de América. Su aprendizaje imparable hará que tome conciencia de sí mismo, ante este evento se intentará su desconexión, pero actuando en defensa propia lanzará un ataque nuclear que desencadenará un “holocausto nuclear” que acabara la inmensa mayoría de vidas humanas.

La postura ética que más se ajusta a la narrativa en este caso (ver anexo II) corresponde a la empática, ya que, mediante la reprogramación de un robot diseñado para eliminar humanos, se consigue proporcionar al protagonista (John Connor) de un prototipo no solo capaz de cumplir las necesidades para su defensa, sino de aprender de él y ayudarle en las misiones que se le requieren durante toda la ficción.



*Ilustración 10. Fotograma de la película “Yo, Robot”, dirigida por Alex Proyas. Fuente: Davis Entertainment Overbrook Entertainment*

#### 4.4.4 Caso 4: Sueños de Robot, Isaac Asimov (1947)

En la ficción que nos expone Asimov se representan dos principales dilemas relativos al futuro de la inteligencia artificial y más concretamente a los AWS. En primer lugar, el prototipo presenta un estatus moral debido a que muestra conciencia de lo que le ocurre, así como de los pensamientos de actuar en su propio beneficio, rigiéndose únicamente por una ley: “Un robot debe proteger su propia existencia” y eliminando parte de lo postulado por la tercera ley de Asimov: “siempre y cuando dicha protección no entorpezca el cumplimiento de la primera y segunda ley”.

*“Estatus moral = conciencia moral del propio derecho, permitiendo la realización de acciones para el bien propio.”*

*Frances M. Kamm (2007)*

Según Bostrom y Yudkowsky (2011), en la actualidad es ampliamente aceptado tanto por la comunidad científica como la filosófica y jurídica que los sistemas de IA no disfrutan de estatus moral. Debido a lo cual, únicamente se fijan límites en caso de que puedan afectar a los derechos de un ser humano, que influiría negativamente a la consideración moral del empleo de AWS.

En segundo lugar, la narrativa de la ficción expone el concepto de “explosión de inteligencia” (Good, 1966). Este hace referencia al futuro hipotético en el cual, los sistemas de inteligencia artificial sean conscientes de su propia existencia y por consiguiente que presenten la capacidad de rediseñarse y crear otros. De esta forma, se expone el desafío ético de futuras “superinteligencias”. Relacionando esta hipótesis circular con la naturaleza variable de los patrones morales, se llega a la conclusión que un sistema autónomo programado para actuar acorde a unos patrones morales (por ejemplo, un AWS programado para evitar dañar a infraestructuras civiles), podrá reajustarlos según varíen conforme a la ética de las personas a lo largo del tiempo. Siguiendo la clasificación del Anexo II, la ficción se aborda desde una postura simpática por la que, debido a que al no cumplirse los requisitos de seguridad ni de necesidad de las capacidades que ofrece el prototipo, este es eliminado.

A continuación, se muestra en forma de tabla resumen el proceso de análisis SFP desarrollado anteriormente (ver tabla 2)



FASE	CASO 1	CASO 2	CASO 3	CASO 4
<b>DESCRIPCIÓN DEL MUNDO, EXPLICACIÓN DEL PROTOTIPO Y PRESENTACIÓN DE LA CIENCIA.</b>	UCAV autónomo con capacidades furtivas. Arquetipos de crecimiento y decadencia, amenazas y nuevas esperanzas	Robot autónomo, regido por leyes menos estrictas y con una inteligencia artificial más avanzada. Arquetipo predominante de poderes fácticos.	Enfrentamiento entre un robot programado por humanos y otro por una inteligencia artificial. Arquetipos de inversión, crecimiento y decadencia, amenazas y nuevas esperanzas.	Robot con un diseño novedoso, similar al cerebro humano, capaz de tener conciencia de sí mismo. Arquetipo de crecimiento y decadencia, amenazas y nuevas esperanzas
<b>IDENTIFICAR EL PUNTO DE INFLEXIÓN CIENTÍFICA Y LAS INTERACCIONES ENTRE LA INNOVACIÓN Y LA SOCIEDAD.</b>	Inicialmente ayuda a la resolución de problemas. Tras un error informático, muestra un comportamiento impredecible y peligroso	Previamente hay una aceptación de la sociedad a la introducción de los robots. La toma decisiones sin reglas creará una situación de peligro para la sociedad.	La aparición de la tecnología causa simultáneamente peligro y protección para la vida de los protagonistas y la sociedad.	El descubrimiento de una tecnología de inteligencia robótica muestra los riesgos de su desarrollo, así como del enorme avance que se puede lograr en este ámbito.
<b>IDENTIFICAR EL PUNTO DE INFLEXIÓN HUMANA</b>	El comportamiento impredecible pone en riesgo vidas humanas. La interacción del prototipo con los humanos consigue evitar que la situación escale, devolviendo al UCAV a la normalidad.	Las características distintivas del prototipo permiten a los humanos evitar un sometimiento de la población por parte de una IA.	El robot programado por los humanos es capaz de aprender valores morales y de esta forma evitar daños humanos.	Los humanos, interrogando al prototipo, son capaces de identificar los peligros de las novedosas características del robot. Tomando la decisión final de eliminarle.



FASE	CASO 1	CASO 2	CASO 3	CASO 4
<b>REFLEXIÓN ÉTICA Y LECCIONES APRENDIDAS</b>	<p>El error informático provoca acciones peligrosas. Los conceptos de responsabilidad y de los límites del aprendizaje de las IA cobran gran importancia. La ficción se aborda desde una postura apática, otorgando más importancia al empleo de la nueva tecnología frente a las vidas humanas que pueda afectar.</p>	<p>La capacidad de una IA de operar con libertad y al margen de unas normas morales genera riesgos. Estos, así como el concepto de regulación cobran gran importancia. Este caso aborda la ficción desde una postura simpática, por la que la sociedad acepta las leyes de la robótica.</p>	<p>El aprendizaje de las normas morales por parte del prototipo, introduce el concepto de reconocimiento moral mutuo entre combatientes, que no ocurre entre las máquinas. Este caso se aborda desde una postura empática.</p>	<p>La tenencia de conciencia moral por parte del prototipo, le otorga un estatus moral, concepto que afecta a la utilización de la IA. También se introduce el concepto de superinteligencias y adaptación a la variabilidad de los patrones morales. En este caso se adopta una postura simpática.</p>

Tabla 2. Tabla resumen del análisis SFP en la que se muestran las fases para cada uno de los casos estudiados. Fuente: elaboración propia.



## 4.5 ESTUDIO SEMICUANTITATIVO

Con la finalidad de recabar más información, así como contrastar las conclusiones obtenidas mediante el empleo de la metodología SFP expuesta en el apartado 2.2, se ha realizado un cuestionario administrativo a 5 expertos de diferentes áreas de conocimientos, entre las que se encuentran la política, el derecho, las relaciones internacionales y el ámbito militar. Debido a las características del estudio semicuantitativo, así como del tema de análisis, es de importancia la obtención de información desde un punto de vista multidisciplinar, con el fin de complementar al máximo el análisis ético.

### 4.5.1 RESULTADOS EN LA ESCALA LIKERT

Se han recopilado las puntuaciones dadas por cada uno de los expertos a cada pregunta. El uso de una hoja de cálculo ha permitido estructurar la información y representarla de forma gráfica para su posterior análisis.

Debido al número de encuestados y a las variables que se estudian en cada una de las preguntas de la entrevista, se ha considerado apropiado la representación gráfica de las respuestas mediante un gráfico radial. Se ha elaborado un gráfico para cada una de las preguntas, la puntuación dada a las preguntas quedará reflejado como un punto sobre cada uno de los radios del pentágono, los cuales representan a cada uno de los encuestados.

ENTREVISTADOS	¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?	¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?	¿Afectaría de forma negativa o positiva el uso de estas tecnologías?	¿Entraría riesgos el uso de estas tecnologías?	¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?
<b>Polítólogo</b>	2	4	2	4	1
<b>Derecho Militar</b>	2	5	1	5	1
<b>Periodismo / RRHH</b>	4	4	3	4	2
<b>Teleco / IA</b>	4	4	3	4	3
<b>Militar</b>	4	3	3	2	2

Tabla 3. Recopilación de las puntuaciones en la escala Likert otorgadas por los expertos a las preguntas del cuestionario. Fuente: elaboración propia.

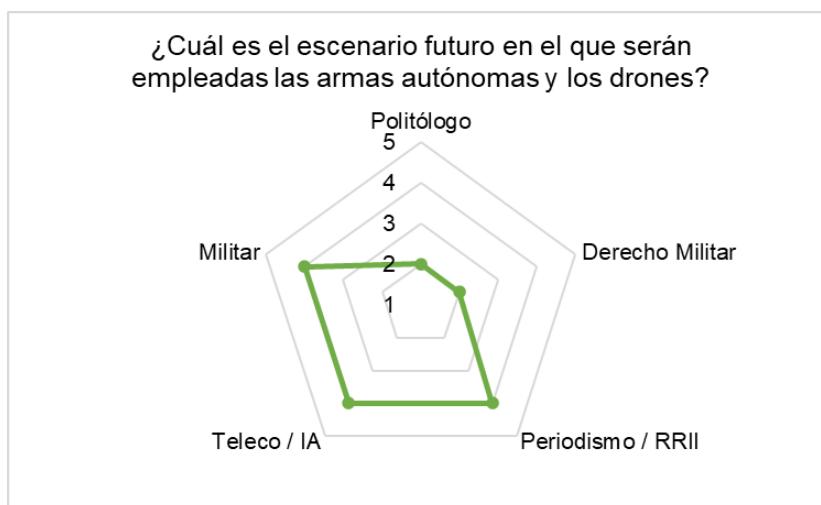
**1ª Pregunta:**



Las respuestas obtenidas de la primera pregunta muestran la falta de consenso acerca de las especulaciones sobre el contexto futuro. El valor más bajo se corresponde con un escenario decadente con apenas regulación al respecto y el valor más alto a una situación controlada en la que el uso de estos sistemas se encuentra regulado y sin riesgos.

Tres de los expertos consideran que el escenario en el que se emplearán las armas y los drones tendrá, en gran medida, regulación al respecto, otorgando una puntuación de 4 sobre el máximo de 5: "se llegarán a acuerdos con el objetivo de evitar desencadenar un posible desastre ante una situación de escalada bélica" (ver anexo VII). En contraposición a esto, las respuestas restantes, por parte del experto en política y derecho militar muestran una posición pesimista acerca de la regulación de estos sistemas, ambos dando una puntuación baja, de 2 sobre 5: "grandes potencias (EE. UU, China y Rusia) han rehuído hasta la fecha la regulación vinculante" (ver anexo VII).

A la vista las puntuaciones (ilustración 11) y las respuestas desarrolladas por los encuestados se concluye que un futuro en el que se actúe al margen de una regulación y otro sujeto a una normativa son igualmente probables.

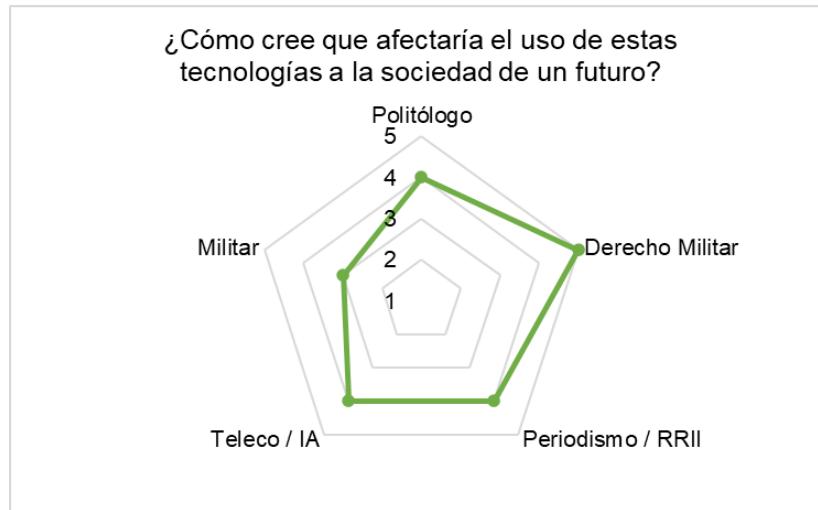


*Ilustración 11 Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca del nivel de regulación. Fuente: elaboración propia.*

## 2ª Pregunta:

Como se puede apreciar en el gráfico, las conclusiones que se pueden extraer de esta pregunta son más claras. El valor más bajo indica un leve impacto y el valor más alto a un impacto alto.

Todos los entrevistados, en mayor o menor medida, comparten la opinión de que el uso de armas autónomas y drones generará un importante impacto en el futuro (ver ilustración 12). Los encuestados están de acuerdo, no solo en un impacto en el ámbito militar sino en el civil debido a la dualidad de aplicación de esta tecnología. "Tendrán un efecto alto en la sociedad porque sus componentes tangibles (sensores, procesadores de alto rendimiento, etc.) y la tecnología intangible (inteligencia artificial) son de uso dual y ya están disponibles en nuestra vida cotidiana. Por tanto, es casi imposible limitar el desarrollo de estas tecnologías" (extraído del anexo VII).

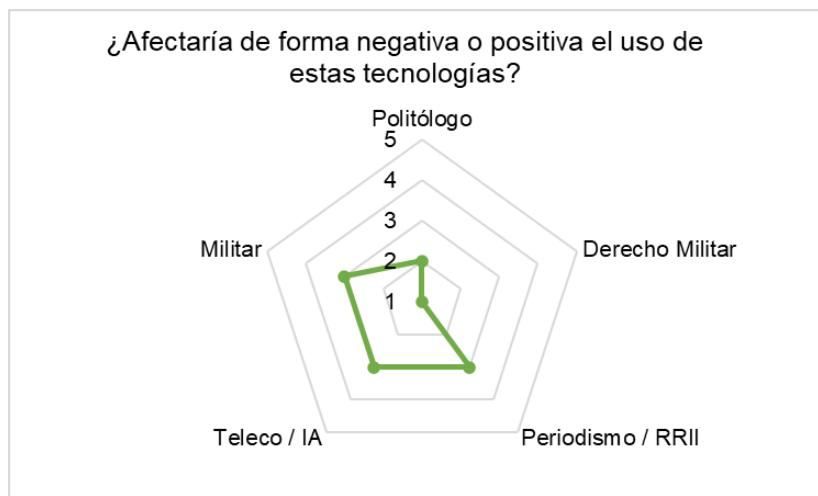


*Ilustración 12. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de la influencia en la sociedad. Fuente: elaboración propia.*

### 3<sup>a</sup> Pregunta:

En la escala Likert empleada en esta pregunta el valor más bajo indica un impacto negativo y el más alto un impacto positivo. Al igual que las respuestas obtenidas en la primera pregunta, los expertos especializados en política y derecho militar eligen una menor puntuación, posicionándose a favor de un impacto negativo acerca del empleo de estos sistemas (ver ilustración 13). El resto de expertos especulan, aunque no de manera clara, sobre el impacto positivo otorgando cada uno de ellos una puntuación de 3 sobre 5: “Más desigualdad por un lado y menor coste en vidas humanas (para los países con estas tecnologías). Posible embrión de más conflictividad por parte de las sociedades sometidas.” (anexo VII).

De esta forma se puede resumir que, aunque presentan mayor prevalencia los efectos positivos sobre los negativos, en los primeros no se obtienen puntuaciones altas, quedando patente la incertidumbre por parte de los expertos acerca de estos y otorgando mayor peso a los negativos por parte de los otros dos encuestados.



*Ilustración 13. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de los efectos producidos. Fuente: elaboración propia.*



#### 4<sup>a</sup> Pregunta:

La escala representa el valor más bajo como un bajo riesgo y el valor más alto como un riesgo alto (ver ilustración 14). Las respuestas obtenidas acerca de los riesgos obtienen, excepto para el encuestado militar, una puntuación alta, por lo que la mayoría de expertos están de acuerdo con los elevados riesgos del empleo de estos sistemas: “El principal riesgo de ellas radica en su capacidad de ser diseñadas para acabar con vidas humanas, el dar elección a una máquina para ejecutar o no un ataque de este tipo entraña un gran riesgo, más todavía cuando no se conocen los posibles errores o problemas de funcionamiento de ellas en diversas situaciones”.

Cabe destacar el contraste de puntuaciones entre esta y la primera pregunta por parte del periodista encuestado, así como del experto en IA, en la que ambos especulan acerca de una situación de regulación, pero al mismo tiempo identifican un alto riesgo futuro.

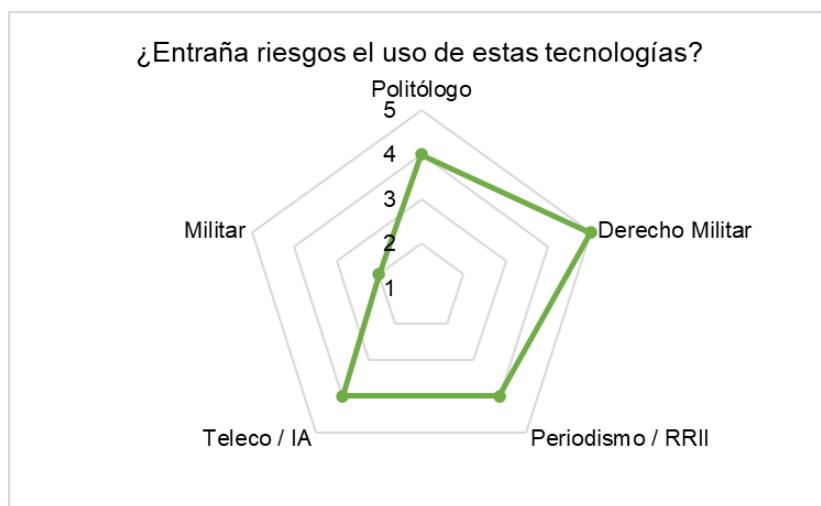


Ilustración 14. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de los riesgos. Fuente: elaboración propia.

#### 5<sup>a</sup> Pregunta:

A la vista de las respuestas obtenidas a la última pregunta, se puede concluir acerca de la poca moralidad en el uso de los sistemas de armas autónomos y los drones. Siendo el valor más bajo moralmente incorrecto y el valor más alto moralmente correcto.

Cuatro de los cinco encuestados otorgan puntuaciones iguales o inferiores a 2, a excepción de estos, el experto en IA opta por una puntuación de 3 sobre 5, volviendo a mostrar una visión más conservadora sobre sus especulaciones (ver ilustración 15). Estos resultados muestran nuevamente la visión pesimista por parte del político y del especializado en derecho militar acerca del empleo de esta tecnología en conflictos bélicos: “un algoritmo no puede actuar como agente moral. Un sistema de armas no puede asumir la responsabilidad de sus actos. No tiene la capacidad cognitiva ni el vínculo emocional que tenemos los seres humanos, aunque este vínculo pueda desaparecer por completo en conflictividades” (extraído del anexo VII).



¿Desde su punto de vista, el uso de armas letales autónomas es moralmente correcto?

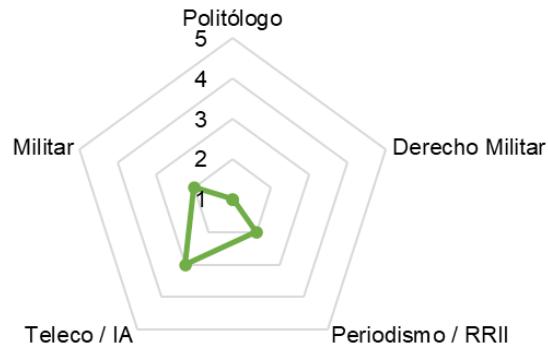


Ilustración 15. Representación en disposición radial de las puntuaciones otorgadas por los expertos acerca de la moralidad. Fuente: elaboración propia.



## 5 CONCLUSIONES

Antes de analizar las conclusiones extraídas de la realización del análisis y del estudio semicuantitativo es necesario aclarar que, debido a las características, así como la variabilidad de los patrones éticos y morales, es de dificultad establecer unas normas éticas que regulen las distintas acciones dirigidas al empleo de sistemas de armas autónomos y drones militares.

Teniendo en consideración esta premisa y con la finalidad de dar una respuesta a los objetivos secundarios planteados, se ha desarrollado de manera extensiva un apartado de antecedentes y marco teórico (apartado 3). En este apartado se han explicado los conceptos de ética y autonomía, así como las aclaraciones acerca de la tecnología de drones militares y armas autónomas, haciendo énfasis en su clasificación por grados de automatización.

En relación con el objetivo principal del trabajo, mediante el análisis de prospectiva basado en SFP, se identifican dos escenarios plausibles acerca del empleo de este tipo de tecnología. En primer lugar, uno con una implementación rápida en el ámbito militar sin apenas regulación previa y con un desconocimiento generalizado de los riesgos derivados de uso, que se puede observar en “*Stealth: La amenaza invisible (2005)*” y “*Terminator 2: el juicio final (1991)*”. Por otro lado, uno con una introducción más gradual, que basa su empleo en un primer momento, fuera del ámbito militar y cuyo uso se realiza de manera extensiva, protagonizado por una mayor regulación al respecto, como se observa en “*Yo, Robot (2004)*” y en “*Sueños de Robot, Isaac Asimov (1947)*”. Contrastando los resultados obtenidos con el estudio semicuantitativo, se observa una similitud con respecto a las especulaciones de los expertos en este asunto: los expertos en política y en derecho militar se posicionan a favor del primer escenario y los expertos en RRII, IA y militar a favor del segundo.

Continuando con el objetivo principal, a lo largo del análisis realizado, se identifican conceptos de importancia ética en relación con los AWS. El primero que se extrae corresponde a la responsabilidad, cualidad que se desvanece cuando se hace uso de estos sistemas. Esto se debe a que, en caso de que se produzcan daños colaterales (como postula uno de los expertos: “no se conocen los posibles errores o problemas de funcionamiento de ellas en diversas situaciones”), se imposibilita la atribución de culpabilidad o castigo al actor de la acción, ya que si la acción es cometida por una máquina, programa informático o IA y éstas no pueden ser castigadas.

Otro concepto que cobra gran importancia y ha sido evaluado en el estudio es el riesgo para la vida. Según lo analizado, se observa un amplio consenso acerca del riesgo que suponen estos sistemas para la vida humana. Este hecho motiva la consideración inmoral y poco beneficiosa acerca de su uso. En relación con la idea anterior se concluye que el empleo de AWS produce una confrontación directa con la dignidad humana provocada por la decisión sobre la vida de un individuo por parte de un programa informático. De este modo el humano pasa de ser un actor en el proceso de decisión a un producto de las mismo. Sobre esta materia los expertos incluidos en el estudio otorgan un mayor peso a los efectos negativos frente a los positivos, reflejando la preocupación acerca del desarrollo que enfrentarán estos sistemas.

Mediante el análisis se identifica la carencia de un reconocimiento moral mutuo en un enfrentamiento protagonizado por AWS o drones. Hecho que supone distinguir el riesgo de atacar o ser atacado con la finalidad de matar, esto impide que la situación arrastre al combatiente, debido a que el adversario tiene ese reconocimiento moral, fenómeno que no ocurriría en un posible conflicto entre sistemas autónomos (situación expuesta por uno de los expertos encuestados como: “nueva modalidad de combate”).

En la última fase de la metodología SFP aplicada, se identifica la idea del estatus moral. En relación a ello, se concluye que los AWS tienen una consideración inmoral debido a la posibilidad de infiligr daño y causar la pérdida de vidas humanas. Para reafirmar la conclusión anterior, los



resultados obtenidos mediante el cuestionario arrojan el consenso por parte de los expertos acerca de la inmoralidad de empleo de los AWS.

## 5.1 LINEAS DE ACCIÓN FUTURA

A la vista de las conclusiones extraídas y con el objetivo de evitar o dar solución a las problemáticas detectadas se proponen a continuación una serie de medidas a implementar para permitir, regular o prohibir el uso en el marco de empleo de los AWS y drones militares. Esta guía proporciona soluciones realistas y adecuadas a los puntos de conflicto evaluados en el apartado anterior:

- Con respecto al valor de la responsabilidad sobre las decisiones se determina que debe permanecer en las personas (comandante, operador, fabricante, programador o estado), de ningún modo debe transferirse totalmente a las máquinas ya que la responsabilidad jurídica aplicable es únicamente humana.
- En relación al riesgo para la vida derivado del empleo de los AWS y los drones militares, se debe realizar un análisis exhaustivo de los riesgos en todo el ciclo de desarrollo de estos sistemas con el fin de evitar pérdidas humanas innecesarias debidas a errores de procesamiento, cálculo, imprecisión o fallos mecánicos.
- Se debe dotar a los sistemas de armas que realicen tareas de manera autónoma, de un código moral actualizado, aplicable a los conflictos e integrado a los algoritmos de decisión que permita cierto reconocimiento moral en enfrentamientos entre hombre y máquina.
- Es imprescindible mantener toda la información relativa al diseño de estos sistemas protegida de apropiaciones indebidas (organizaciones terroristas, crimen organizado, ciber amenazas, etc.).
- Es necesario mantener un equilibrio entre operatividad y seguridad, fijando límites invariables por diseño que impidan que se ejecuten acciones fuera del ámbito para el que fueron creados, impidiendo la reprogramación autónoma del sistema, permitiendo la variación del grado de automatización en cualquier momento de la “cadena de muerte” e implementando mecanismos de seguridad que invaliden el sistema en caso de desvinculación con el operador.



## 6 REFERENCIAS BIBLIOGRÁFICAS

- Aguilar, J., Ter, O. and Morantes, W. (2006) 'Prospectiva Tecnológica'.
- Historia de la Robótica* (2017). Robótica Educativa. Disponible en: <https://www.roboticaeducativa.pe/2017/10/20/historia-de-la-robotica/>. (Consultado el 19-09-2021)
- Asaro, P. M. (2008) 'How just could a robot war be?', *Frontiers in Artificial Intelligence and Applications*, 175(1), pp. 50–64.
- Asimov, I. (2004) 'Sueños de Robot'. Barcelona: Debolsillo, 2<sup>a</sup>edición
- Aznar Fernández Montesinos, F. (2016) 'Los Militares y la Tecnología', *Instituto Español de Estudios Estratégicos*, pp. 1–18. Disponible en: [http://www.ieee.es/Galerias/fichero/docs\\_analisis/2016/DIEEEA72-2016\\_Militares\\_Tecnologia\\_FAFM.pdf](http://www.ieee.es/Galerias/fichero/docs_analisis/2016/DIEEEA72-2016_Militares_Tecnologia_FAFM.pdf). (Consultado el 25-09-2021)
- Baeza, C. (2020) 'Using desing fiction to teach ethics in design', *IDSA*, pp. 1–9.
- Beary, J. F. (1983) 'Department of Defense', *Military Medicine*, 148(11), pp. 857–863. doi: 10.1093/milmed/148.11.857.
- Benavides, L. (2021) 'Proyectos utópicos. Estrategia artística para potenciar la imaginación y la capacidad creadora, por el beneficio de la sociedad', *Arte, Individuo y Sociedad*, 33(3), pp. 819–836. doi: 10.5209/aris.70242.
- Bennett, M. and Johnson, B. D. (2016) 'Dark future precedents: Science fiction, futurism and law', 1(September), pp. 1–8. doi: 10.3233/978-1-61499-690-3-506.
- Bostrom, N. and Yudkowsky, E. (2011) 'The ethics of artificial intelligence', in *Cambridge Handbook of Artificial Intelligence*, pp. 166–168. doi: 10.1016/j.mpmed.2018.12.009.
- Cameron, J. (1991) 'Terminator 2: el juicio final'. *Carolco Pictures, Pacific Western; Lightstorm Entertainment, T2 Productions*.
- Castrillón Gómez, O. D., Rodríguez Córdoba, M. D. P. and Leyton Castaño, J. D. (2008) 'Ética E Inteligencia Artificial ¿Necesidad O Urgencia?', *CISCI 2008 - Septima Conferencia Iberoamericana en Sistema, Cibernetica e Informatica 5to SIECI 2008, 3er Simposium Internacional en Comunicacion del Conocimiento y Conferencias, CCC 2008 - Memorias*, 3, pp. 121–126.
- Col. Gómez de Ágreda, A. (2021) *Ética del ecosistema híbrido cognitivo entre el espacio físico y el ciberespacio. Aproximación desde el caso de la inteligencia artificial*. Tesis. Universidad politécnica de Madrid.
- Colesniuc, D. (2013) 'Cyberspace and Critical Information Infrastructures', *Informática Económica*, 17(4/2013), pp. 123–132. doi: 10.12948/issn14531305/17.4.2013.11.
- Curtis E. Lemay Center (2019) 'Dynamic targeting and the tasking process', *Air Force Doctrine Publication (AFDP)*, (March), pp. 43–47.
- David Johnson, B. (2012) 'Science Fiction Prototyping: Designing the Future with Science Fiction Designing the Future with Science Fiction', *Morgan & Claypool All*, p. 32.
- Fernani, A. and Song, Z. (2020) 'The six scenario archetypes framework: A systematic investigation of science fiction films set in the future', *Futures*, 124(May), p. 102645. doi: 10.1016/j.futures.2020.102645.
- Gomez, P. (2021) *Origen e historia de la robótica*. Disponible en: <https://www.hipernexo.com/robotica/historia-de-la-robotica/>. (Consultado el 15-09-2021)



González Estévez, A. E. (2019) *Robótica, Inteligencia Artificial, Vehículos autónomos y Drones: Responsabilidad civil y seguro*, Escuela Técnica Superior de Ingeniería de Sistemas Informáticos. Disponible en: <http://msde.etsisi.upm.es/index.php/project/robotica-inteligencia-artificial-vehiculos-autonomos-y-drones-responsabilidad-civil-y-seguro/>. (Consultado el 13-10-2021)

Good, I. J. (1966) 'Speculations Concerning the First Ultraintelligent Machine' The first draft of this monograph was completed in April 1963, and the present slightly amended version in May 1964. I am much indebted to Mrs. Euthie Anthony of IDA for the arduous task of typin', *Advances in Computers*, 6(C), pp. 31–88. doi: 10.1016/S0065-2458(08)60418-0.

Haider, A. and Catarrasi, B. (2016) 'Future Unmanned System Technologies: Legal and Ethical Implications of Increasing Automation'.

Jensen, T. and Vistisen, P. (2017) 'Ethical Design Fiction', *The ORBIT Journal*, 1(2), pp. 1–14. doi: 10.29297/orbit.v1i2.56.

Kant, I. and Aramayo, R. R. (Trad) (2012) 'Immanuel Kant Fundamentación para una metafísica de las costumbres. Versión castellana y estudio preliminar H Alianza. Editorial El libro de bolsillo', p. 219.

Lind, W. S. (2004) *Understanding Fourth Generation War*, original.antiwar.com. Disponible en: <https://original.antiwar.com/lind/2004/01/15/understanding-fourth-generation-war/>. (Consultado el 12-09-2021)

Lt. Col. Leslie Pratt. (2008) 'MQ-1 Predator'. U.S. Air Force Photo.

Majumdar Roy Choudhury, L. (2021) *Carta de fecha 8 de marzo de 2021 dirigida a la Presidencia del Consejo de Seguridad por el Grupo de Expertos sobre Libia establecido en virtud de la resolución 1973*.

Monasterio Astobiza, A. (2017) 'Ética militar y robótica: Drones, robots militares, sistemas de armas autónomas letales y la transformación de la naturaleza de la guerra en el siglo XXI', *Actas del II Congreso internacional de la Red española de Filosofía. Las fronteras de la humanidad*, I, pp. 167–180.

Nagel, T. (2012) 'War and Massacre', *Philosophy & Public Affairs*, Vol. 1, No. 2 (Winter, 1972), pp. 123-144, 1(2), p. 32.

Pascual, D. (2017) 'Inteligencia Artificial : Un Panorama De Algunos De Sus Desafíos Éticos Y Jurídicos', p. 56.

Proyas, A. (2004) *Yo, Robot*. Davis Entertainment Overbrook Entertainment

Relay, I. C. (2019) ¿UAV, UAS, RPAS o drones ?, *Instituto Nacional de Técnica Aeroespacial, INTA*

Robert Cohen (2005) "Stealth La amenaza invisible". Gaumont Columbia Tristar Films.

Rood, G. (2020) 'The Royal Aircraft Establishment Farnborough 100 years of Innovative Research, Development and Application', pp. 38–98.

Ross, L. y (2018) *Asimov's 3 laws of robotics, updated for the drone age*, Fast Company.

Scharre, S. (2016) 'Autonomous Weapons and Operational Risk Paul Scharre Senior Fellow and Director of the 20YY Future of Warfare Initiative Center for a New American Security'

Simpson, T. W. and Müller, V. C. (2016) 'Just War and Robots' Killings', *Philosophical Quarterly*, 66(263), pp. 302–322. doi: 10.1093/pq/pqv075.

Sparrow, R. (2007) 'Killer robots', *Journal of Applied Philosophy*, 24(1), pp. 62–77. doi: 10.1111/j.1468-5930.2007.00346.x.

Sterling, B. (2013) *Patently untrue*, [www.wired.co.uk](http://www.wired.co.uk). Disponible en:



- <https://www.wired.co.uk/article/patently-untrue>. (Consultado el 23-09-2021)
- Vanatta, N. and Johnson, B. D. (2019) 'Threatcasting: a framework and process to model future operating environments', *Journal of Defense Modeling and Simulation*, 16(1), pp. 79–88. doi: 10.1177/1548512918806385.
- Villavicencio Mavrich, H. (2005) 'Tecnología de futuro: Cirugía robótica Da Vinci', *Actas Urologicas Espanolas*, 29(10), pp. 919–921. doi: 10.1016/S0210-4806(05)73370-2.
- Wendell W (2009) *Moral Machines: Teaching Robots Right from Wrong*.
- Westhues, A. (2020) *Sistemas de armas autónomas letales : ¿Autónomas o Automatizadas?*, Trabajo de Fin de Master. Instituto Universitario General Gutierrez Mellado.



## ANEXOS



## Anexo I. SFP

### Paso 1: Elegir la ciencia y construir el mundo

Para ello, se debe elegir la tecnología, ciencia o problema que se quiere explorar mediante el prototipo. Se debe empezar considerando las versiones futuras de la tecnología que se ha elegido explorar, para las que se pueden formular varias preguntas sencillas:

- ¿Qué implicaciones tendría el uso en masa de esta tecnología?
- ¿Qué es lo que podría ir mal y cómo afectaría a las personas y contexto de la historia?
- ¿Cuál es lo mejor que podría ocurrir y como mejoraría el contexto y las vidas de las personas en la historia?
- ¿Si esta tecnología estuviera accesible a todo el mundo como funcionaría?

Una vez se hayan obtenido las respuestas a estas preguntas, se puede comenzar con una lluvia de ideas. Hay que recordar que se está ubicando la idea o el tema a tratar en un contexto real, pero a la hora de realizar el análisis, a pesar de que esté situado en un futuro lejano, el mundo que se pretende mostrar debe sentirse real, es decir, todavía debe regirse por las leyes y lógica científicas. Además, deben aparecer personas reales y el situar el escenario en un futuro cercano, en un mundo que sea conocido y que sea útil para explorar y conocer mejor los efectos de esta tecnología, es esencial.

### Paso 2: Identificar el punto de inflexión científica

Una vez se ha elegido la idea es hora de ver que ocurre cuando la tecnología que se quiere analizar entra a funcionar el contexto creado anteriormente. Para ello, es importante centrarse en las personas y sistemas en tu mundo, de esta forma se estudian los efectos que tiene la nueva tecnología en la vida diaria, política, economía y en el caso de este trabajo, en la ética.

### Paso 3: Considerar las interacciones recíprocas de la innovación y la sociedad

Una vez se libere la idea o tecnología en el mundo que se ha creado, ésta tendrá efectos en cadena. La historia cambiará y se adaptará a la tecnología que se ha introducido. En esta situación tenemos que hacernos otras preguntas:

- ¿La introducción de esta tecnología ha mejorado sus vidas?
- ¿La introducción de esta tecnología ha empeorado sus vidas?
- ¿Se han adaptado a los problemas que han surgido?
- ¿Se han adaptado a las ventajas u oportunidades que ha traído?

Es en este punto donde se debe prestar atención a la trama que se desarrolla en la película, comic, o historia de ciencia ficción. En algunas situaciones se observará como la situación se volverá crítica, hasta tal punto que las personas deban actuar. Esto no es sólo una narración de la historia, sino que también es beneficioso para estudiar cómo se desarrolla la tecnología o ciencia que se ha elegido. El llevar la trama a los extremos positivos o negativos, expondrá nuevas áreas de exploración y ayudará a aclarar la idea.

### Paso 4: Identificar el punto de inflexión humana.

Después de explorar los extremos de la situación en la que se encuentra el prototipo de ciencia ficción, las personas que aparecen en la trama pueden estar en peligro o no:

- ¿Qué necesitan hacer las personas para sobrevivir?
- ¿Qué ramas de la ciencia se han seleccionado y que situación ha provocado?

En este punto, se observa qué harán los personajes para adaptarse a la ciencia y tecnología introducidos en el paso 2. Para ello, las personas o la ciencia cambiarán para adaptarse; ambos



resultados pueden estar en consonancia en la historia que se ha creado. Es vital que los cambios sean plausibles para el mundo virtual y se mantengan dentro de los límites de la ciencia. Cuanto más se acerque el límite de la ciencia ficción al de la ciencia real, más valioso será la información obtenida mediante el empleo de SFP.

#### **Paso 5: Reflexionar sobre lo aprendido**

Después del paso 4 debemos hacernos las siguientes preguntas:

- ¿Qué se ha aprendido al colocar la idea o tecnología en el mundo que se ha creado?
- ¿Cómo ha cambiado el mundo?
- ¿Cómo han cambiado las personas, la sociedad y los sistemas?
- ¿Qué elementos se podrían modificar?
- ¿Qué precauciones se deben de tener en cuenta?
- ¿Qué debería ser diferente?
- ¿Qué debería permanecer igual?
- ¿Qué se ha descubierto utilizando SFP?

Los cinco pasos expuestos anteriormente constituyen una guía para examinar la ciencia, así como el impacto que tendrá. Al igual que ocurre en muchos procesos, una vez se esté familiarizado, se pueden realizar modificaciones, pero estos 5 pasos forman un esquema sólido para analizar una historia de ciencia ficción, basada específicamente en hechos científicos.



## Anexo II. Ficción de diseño ético

### Paso 1: Relato o explicación

Presenta el mundo a la audiencia, incluye decisiones del diseñador sobre cómo verse el mundo y que partes son importantes. Siguiendo el modelo de Auger sobre los futuros especulativos, el escenario presentado a través de la historia muestra posibles presentes alternativos o futuros perdidos. Enmarcando el presente real con el pasado, permite a la audiencia reconocer tecnologías emergentes y el tiempo del proceso para que sean domesticadas.

### Paso 2. Prototipo

El prototipo es el diseño propuesto desde un concepto hasta un prototipo funcional. Este solo existe dentro del contexto narrativo de la historia, además de existir como uno funcional, podría tratarse de un objeto de I+D en el mundo real. Este debe funcionar como se dice en la narrativa y no como lo hace actualmente, para ello debe de encajar en el mundo de la historia y debe de ser su punto de apoyo.

### Paso 3. Discurso

El modelo de Auger muestra la importancia del proceso de domesticación de las tecnologías emergentes y el encuadre de los presentes alternativos. Este desafío está cubierto parcialmente por el discurso que rodea al diseño. Las manifestaciones reales como implementadas, así como las propuestas y contadas a través de la narrativa forman un “espacio de diseño”, dentro del cual pueden existir múltiples conceptos para una nueva propuesta de diseño. El autor debe basarse en estructuras existentes con la finalidad de que la audiencia comprenda el concepto, dirigiendo su atención a los problemas del diseño.

### Paso 4. Posturas éticas

Este apartado a diferencia del resto puede estar fuera del alcance de la ficción de diseño. En muchas ocasiones, ésta puede querer mostrar el nuevo diseño desde un punto de vista positivo o neutral, obviando cualquier desafío ético. Si la ficción mantiene una postura no vinculante con las cuestiones éticas, la funcionalidad y utilidad tendrán un mayor protagonismo respecto a la experiencia o participación del usuario. En este caso, la audiencia necesita mostrar sus propias conclusiones éticas. Según Jensen and Vistisen, (2017) la ficción de diseño debe ser abordado como una narrativa que abre a la discusión, favorece la inmersión de la audiencia y de esta forma permite que ésta entienda las posibilidades y desafíos futuros.

### Paso 5. Desarrollo del carácter

Al igual que las cuestiones éticas, el elemento que necesita una postura empática es el desarrollo del carácter. En la ficción de diseño se explica el futuro de un dispositivo y sus implicaciones para las personas afectadas por ella. En cuanto a este elemento de la ficción, debe usarse con la finalidad de que el personaje crezca y se desarrolle

### Paso 6. Apelación retórica

Este elemento es necesario para conmover al público e involucrar a la audiencia, con el fin de favorecer el diseño. Para ello el pathos es esencial para mantener una actitud comprensiva hacia el diseño. Si la ficción de diseño necesita explorar problemas éticos importantes, la audiencia tiene que confiar en el narrador, los personajes y el mundo presentados. El narrador debe respetar a la audiencia para mostrar desafíos y posibilidades, así como problemas y soluciones. Se utilizan los logotipos como elementos retóricos de manera que sirven para centrarse en el prototipo en lugar del impacto a los usuarios. Los logotipos deben de ser el atractivo de la ficción de diseño, los cuales devuelven la atención al estado actual.



A través de la elección deliberada de las posturas éticas y retóricas, la ficción de diseño cambiará su enfoque en el que se muestra a la audiencia. Por ejemplo, una postura apática se centra en la coherencia y ontología del mundo, una simpática, en el comportamiento del prototipo dentro del mundo; por último, una postura empática analiza de forma crítica la utilidad, usabilidad y conveniencia del espacio discursivo creado alrededor del prototipo, así como la implicación ontológica y epistemológica del artefacto en el mundo y como afecta al usuario. Las elecciones deliberadas por parte del diseñador son necesarias para la maduración de la ficción de diseño y como herramienta para discutir la estrategia futura de la tecnología elegida. De esta forma, la narración y posturas éticas e intenciones del diseñador deben de adaptarse a la audiencia y prototipo en cuestión.

Respecto al marco de ficción de diseño ético, las teorías y métodos dentro de una narrativa, literatura y retórica pueden añadir comprensión y creación de prototipos para el proceso de diseño. La narrativa ética permite la creación de un discurso en el que los desafíos y problemas pueden abordarse junto con las responsabilidades del diseñador. De esta forma, creando un mundo de historias en el que el prototipo cobra vida y produce un impacto en las personas, obtenemos un medio en el que se puede mostrar y evaluar.

La ficción no solo tiene como objetivo proponer una ontología de "lo que podría ser el mundo" o una epistemología de "cómo la gente podría percibir el mundo", sino más bien desafiar activamente la audiencia a cuestionar tanto la ontología como la epistemología del mundo de la historia y su prototipo. De esta forma se orienta que postura ética se adopta, entre apática, comprensiva o empática. Tomando como base las discusiones anteriores, Jensen y Vistisen proponen un marco de análisis que se puede leer en vertical o en horizontal dependiendo de si se quiere utilizar como herramienta de análisis o de síntesis respectivamente (tabla 4).

DISEÑO DE FICCIÓN			
NARRATIVA	Logos + pathos	Pathos + problemas éticos menores y desarrollo del carácter	Ethos + pathos + problemas éticos menores y desarrollo del carácter
CONSTRUCCIÓN DEL MUNDO	Consistencia del discurso (logos)	Ayuda en el discurso (pathos)	Responsabilidad en el discurso (ethos)
	Ontología (Este es el mundo)	Epistemología (Así es como actuamos en el mundo)	Cuestionarse la ontología y epistemología (¿Cómo debemos actuar? ¿Qué mundo queremos?)
POSTURAS ÉTICAS	Apatía (Sistema>usuario)	Simpatía (otorgar a los usuarios lo que quieren)	Empatía (otorgar a los usuarios lo que necesitan)

Tabla 4. Matriz de ficción de diseño. Fuente: Jensen and Vistisen (2017).



### Anexo III. Tabla resumen evaluación del objetivo del diseño

ÉTICA DE LA VIRTUD	DEONTOLOGÍA (ETICA DEL DEBER)	CONSECUENCIALISMO
<p>Trata con lo correcto y lo incorrecto sobre acciones individuales. Es una guía sobre una serie de características y comportamientos que una buena persona espera conseguir.</p>	<p>La ética basada en el deber enseña que algunos actos son correctos o incorrectos debido al tipo de cosas que son, y las personas tienen el deber de actuar en consecuencia, independientemente de las buenas o malas consecuencias que puedan producirse.</p>	<p>Lo correcto o lo incorrecto depende de las consecuencias de un acto, conforme mejores consecuencias ocurran, mejor es el acto.</p>
<p><b>¿Es el diseño moralmente correcto? En caso negativo, ¿por qué no? ¿Cómo se podría cambiar?</b></p>		

Tabla 5. Evaluación del objetivo del diseño. Fuente: Baeza (2020).



## Anexo IV. Cuestionario Administrativo a expertos

### **1. ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?**

Elegir un valor entre una escala del 1 al 5, haciendo referencia el valor más bajo a un escenario decadente con apenas regulación al respecto y el valor más alto a una situación controlada en la que el uso de estos sistemas se encuentra regulado y sin riesgos.

Explique brevemente su respuesta:

### **2. ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?**

Elegir un valor entre una escala del 1 al 5, el valor más bajo indica un leve impacto y el valor más alto a un impacto alto.

Explique brevemente su respuesta:

### **3. ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?**

Elegir un valor entre una escala del 1 al 5, el valor más bajo indica un impacto negativo y el valor más alto un impacto positivo.

Explique brevemente su respuesta:

### **4. ¿Entraña riesgos el uso de estas tecnologías?**

Elegir un valor entre una escala del 1 al 5, el valor más bajo indica un bajo riesgo y el valor más alto un alto riesgo.

Explique brevemente su respuesta:

### **5. ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?**

Elegir un valor entre una escala del 1 al 5, siendo el valor más bajo moralmente incorrecto y el valor más alto moralmente correcto.

Explique brevemente su respuesta:



## Anexo V. Listado de obras de ciencia ficción

1. The Matrix (1999)
2. RoboCop (1987)
3. Star Wars (1977)
4. Máquinas como yo, Ian McEwan (2019)
5. Iron Man 2 (2010)
6. AI. Artificial Intelligence (2001)
7. Yo, Robot. Isaac Asimov (1950)
8. Dune, Frank Herbert (1965)
9. Stealth: La amenaza invisible (2005)
10. Yo, Robot (2004)
11. Terminator 2: el juicio final (1991)
12. Sueños de Robot, Isaac Asimov (1947)



## Anexo VI. Modelo de los 6 Arquetipos

### 1. Crecimiento y decadencia

Este arquetipo implica la continuación del statu quo capitalista actual, que crece aún más desenfrenadamente. Las corporaciones reinan sin mezcla, extendiendo potencialmente su poder a la vigilancia, la seguridad urbana, la gestión de infraestructuras públicas y la aplicación de la ley. El poder gubernamental está ausente o marginado. Las tecnologías actuales también crecen constantemente, impulsadas por ganancias monetarias y controladas por corporaciones.

En la sociedad se encuentran indicios de colapso o decadencia social. La decadencia puede manifestarse de varias formas, desde condiciones de vida abyectas y disparidades patentes hasta la mala gestión de la higiene urbana y la desolación del entorno atmosférico, o en una combinación de estos elementos. La decadencia también puede manifestarse en la decadencia de valores comunes, o en el conflicto entre valores entre diferentes grupos de individuos/facciones. Por ejemplo, el liderazgo puede ser malvado, principalmente en busca de ganancias y dominante, mientras que los subordinados están subyugados.

### 2. Amenazas y nuevas esperanzas

En este arquetipo, ningún cambio significativo afecta a la humanidad, y las condiciones de vida humana son muy similares a las actuales. Sin embargo, un evento o fenómeno catastrófico o apocalíptico inminente amenaza la existencia de la humanidad. Este acontecimiento inminente puede tomar varias formas, incluidos los desastres ambientales, las destrucciones provocadas por el hombre o la invasión de extraterrestres.

Organismos gubernamentales u organizaciones militares nacionales y supranacionales colaboran para diseñar un plan global de rescate, mientras que el sector privado es menos relevante. Los individuos sacrifican su riqueza personal, sus afectos e incluso sus vidas por el bien común para salvar al mundo mientras la humanidad se une para luchar contra un enemigo común.

### 3. Mundos baldíos

En este arquetipo, ya ha ocurrido un evento o fenómeno catastrófico, trayendo consigo transformaciones sustanciales a escala global. El entorno atmosférico a menudo se ve afectado de manera perniciosa, lo que obliga a los humanos a adaptarse a condiciones de vida drásticas.

Muchas veces, en el contexto de una severa escasez de recursos, la civilización humana ha retrocedido al nivel de sustento. La economía de mercado ha dado paso a sistemas económicos más rudimentarios, como el trueque o el uso del agua, el petróleo o la arena como moneda. Pocos supervivientes viven en comunidades tribales dispersas, luchando por sobrevivir y, a menudo, explotados por bandas de forajidos. Los líderes locales tiránicos a menudo subyugan a estas comunidades indefensas, expropiando sus recursos. Los individuos luchan entre sí por la supervivencia. Otras veces, los humanos abandonan la tierra por completo debido a condiciones ambientales inhabitables, siendo esta una variación del arquetipo, aunque con las mismas premisas iniciales.

### 4. Poderes fácticos

En este arquetipo, ya ha ocurrido un evento o fenómeno catastrófico, a menudo provocado por el hombre. Aunque esto ha dejado una cicatriz en la especie humana hasta el punto de que la población a menudo se reduce significativamente, la humanidad retoma su camino hacia el progreso rápidamente a partir de entonces. Sin embargo, estrictos poderes totalitarios o dictatoriales emergen de este pasado accidentado, aparentemente para prevenir cuidadosamente la ocurrencia de otros eventos o fenómenos devastadores provocados por el hombre en el futuro. La tecnología es avanzada, pero centralizada en manos de organismos gubernamentales y utilizada como instrumento de control. En paralelo al progreso técnico, los



derechos, la felicidad, la libertad y las emociones de los ciudadanos están limitados desde arriba. Los individuos intentan emanciparse, rebelarse contra los regímenes existentes o desarraigárselos.

## 5. Desorden

En este arquetipo, aunque en ausencia de aparentes cambios transformacionales en la economía o el ambiente atmosférico, la humanidad enfrenta problemas endógenos estructurales. El mundo está plagado de cualquiera de los siguientes: crimen endémico, malestar social y desorden, pobreza generalizada, ignorancia, infertilidad, confrontación violenta y guerra, hambrunas o pandemias; o por una combinación de estos. Aunque el sector privado todavía está presente, las organizaciones militares y policiales, ya sean oficiales o no oficiales, tienen un papel más central en este futuro. Los esfuerzos individuales se concentran en restaurar o mantener la justicia, el orden o la protección de los ciudadanos.

## 6. Inversión

En este arquetipo, el papel de la humanidad se invierte, ya que es superado o subyugado por una civilización, agente u organismo superior. Los seres humanos ya no dominan el planeta. A menudo, en cambio, están dominados por criaturas de mayor destreza física, de las que se convierten en presas. Las especies alienígenas que invaden el planeta o la galaxia entera son un ejemplo, ya sea de apariencia monstruosa o antropomórfica. Sin embargo, esta entidad superior también podría manifestarse de maneras más sutiles, como un creador o supervisor ostensible con quien la humanidad no debería interferir.



Ilustración 16. Representación de los 6 arquetipos. Fuente: Fergnani and Song (2020).



## Anexo VII. Respuestas al cuestionario administrativo

### Primer encuestado

<b>Formación / Cargo / Área de conocimiento</b>	Graduado en Ciencias Políticas y de la Administración. Máster en Paz, Seguridad y Defensa.
<b>Fecha</b>	07/11/2021

*Tabla 6. Perfil del primer encuestado. Fuente: elaboración propia*

**1ª Pregunta:** ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?

**Respuesta:** Veo un escenario con pocas opciones de una regulación internacional, porque las grandes potencias (EE. UU, China y Rusia) han rehusado hasta la fecha la regulación vinculante, bien rechazándola directamente (Rusia y EE. UU) o rechazando solamente el empleo de armas con plena autonomía, pero sin rechazar su desarrollo (China). Creo que la competencia geopolítica se desarrolla aquí solamente en una dimensión más.

**2ª Pregunta:** ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?

**Respuesta:** Estas tecnologías tendrán un efecto alto en la sociedad porque sus componentes tangibles (sensores, procesadores de alto rendimiento, etc.) y la tecnología intangible (inteligencia artificial) son de uso dual y ya están disponibles en nuestra vida cotidiana. Por tanto, es casi imposible limitar el desarrollo de estas tecnologías. Por otra parte, es posible que una parte de los conflictos se lleven a cabo usando estas tecnologías porque reducen el riesgo personal de las partes implicadas. Esto ocurriría especialmente en los conflictos internos. Lo que no creo es que provoquen rechazo total en la población como en el caso de las armas nucleares durante la Guerra Fría.

**3ª Pregunta:** ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?

**Respuesta:** Depende de a quién ... a empresas desarrolladoras de forma positiva porque es un negocio lucrativo que está en pleno auge. A países con potencial agresivo y sin grandes preocupaciones democráticas también les afectaría de forma positiva porque estos sistemas de armas podrían reducir su propio riesgo en el conflicto. También son un componente pequeño que puede usarse perfectamente para ataques muy puntuales, como a las refinerías Saudí Aramco. Al ser limitados, se podrían incorporar en acciones de guerra híbrida. Pero, encima de todo, el mayor impacto negativo es para la población civil, por esto mi puntuación es baja.

**4ª Pregunta:** ¿Entraña riesgos el uso de estas tecnologías?

**Respuesta:** Creo que el riesgo es alto, porque las armas autónomas pueden bajar el umbral de iniciar a un conflicto o responder de forma precipitada a una agresión. También es posible la desencadenación "automática" de un conflicto y el incremento automático/autónomo de las acciones bélicas cuando dos sistemas se refuerzan mutuamente en una espiral descontrolada. Otros riesgos son la carrera armamentística entre las grandes potencias y la proliferación de estas a otros países o actores no estatales.

**5ª Pregunta:** ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?

**Respuesta:** Totalmente incorrecto porque un algoritmo no puede actuar como agente moral. Un sistema de armas no puede asumir la responsabilidad de sus actos. No tiene la capacidad



cognitiva ni el vínculo emocional que tenemos los seres humanos, aunque este vínculo pueda desaparecer por completo en conflictividades.

### **Segundo encuestado**

<b>Formación / Cargo / Área de conocimiento</b>	Graduado en derecho, especializado en Derecho Militar
<b>Fecha</b>	25/11/2021

*Tabla 7. Perfil del segundo encuestado. Fuente: elaboración propia*

**1ª Pregunta:** ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?

**Respuesta:** A pesar de la internacionalidad de los protocolos adicionales I y II y los Convenios de Ginebra que se pueden aplicar a las armas autónomas y a los drones, para muchos países como EEUU o Pakistán no tienen carácter vinculante. Debido a esto, veo poco probable un consenso internacional acerca de una regulación que sea reconocida por todos los países que hagan uso de estos.

**2ª Pregunta:** ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?

**Respuesta:** Tendrán un gran impacto, tanto en el ámbito militar como en el civil. Sin ninguna duda afectará su uso en ámbitos tanto económicos, políticos y sociales. Su empleo en la toma de decisiones y la sustitución de la intervención humana serán claves en un futuro a medio plazo.

**3ª Pregunta:** ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?

**Respuesta:** En mi opinión, el impacto generado de su uso será negativo, principalmente porque el valor de la vida o de la actividad humana perderá importancia. Por ejemplo, en un conflicto con una gran brecha tecnológica entre dos bandos, o en el ámbito civil la sustitución de la mano de obra humana por máquinas cada vez más automatizadas.

**4ª Pregunta:** ¿Entraña riesgos el uso de estas tecnologías?

**Respuesta:** Al igual que los descubrimientos en materia nuclear, química y biológica, las armas autónomas pueden generar nuevos riesgos para la vida humana en caso de que un conflicto a gran escala se desencadene. Pero el principal riesgo de estas es que sean usadas por organizaciones terroristas.

**5ª Pregunta:** ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?

**Respuesta:** El principal problema moral y ético de estos sistemas es su capacidad de decisión sobre las vidas humanas, desde mi punto de vista esto no es moral ya que evade los principios del uso legítimo de la fuerza.



### Tercer encuestado

<b>Formación / Cargo / Área de conocimiento</b>	Licenciado en Periodismo, experto en Inteligencia, Diplomacia y Relaciones Internacionales
<b>Fecha</b>	08/12/2021

*Tabla 8. Perfil del tercer encuestado. Fuente: elaboración propia*

**1ª Pregunta:** ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?

**Respuesta:** Apuesto por un futuro con un amplio marco normativo en todo lo relacionado con estos sistemas, debido a las situaciones de riesgo que pueden generar para la vida humana. Al igual que con el armamento nuclear, cuando las armas autónomas comiencen a convertirse en los principales instrumentos de disuasión, las grandes potencias llegarán a acuerdos con el objetivo de evitar desencadenar un posible desastre ante una situación de escalada bélica.

**2ª Pregunta:** ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?

**Respuesta:** Producirán un gran impacto en todos los ámbitos, pero en el de la defensa, su uso introducirá nuevas modalidades de combate, como las que se están viendo en conflictos actuales en los que se hacen uso de drones. La importancia del coste de vidas humanas disminuirá y los sistemas de armas de este estilo ocuparán su lugar.

**3ª Pregunta:** ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?

**Respuesta:** Mi elección en esta pregunta no es clara, ya que como he explicado en la anterior, su empleo en otros ámbitos a parte del bélico puede aportar efectos positivos. Por el contrario, el uso de estas armas con la finalidad de acabar con vidas humanas podría provocar problemas, por ejemplo, si estos sistemas acaban en manos de grupos terroristas o en naciones que no acepten una futura normativa relativa a su empleo.

**4ª Pregunta:** ¿Entraña riesgos el uso de estas tecnologías?

**Respuesta:** El principal riesgo de ellas radica en su capacidad de ser diseñadas para acabar con vidas humanas, el dar elección a una máquina para ejecutar o no un ataque de este tipo entraña un gran riesgo, más todavía cuando no se conocen los posibles errores o problemas de funcionamiento de ellas en diversas situaciones.

**5ª Pregunta:** ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?

**Respuesta:** Consideró que su uso, en la mayoría de situaciones es poco correcto por el daño que puedan causar, por ejemplo, decidiendo sobre vidas humanas. A pesar de ello, no lo considero moralmente incorrecto ya que la nación que las emplea está evitando que caigan muchos de sus militares en combate. Debido a esto su uso puede clasificarse de distintas formas según la perspectiva que se tome.



## Cuarto encuestado

<b>Formación / Cargo / Área de conocimiento</b>	Graduado en Ingeniería en Tecnologías de Telecomunicaciones, master en Inteligencia Artificial
<b>Fecha</b>	27/12/2021

*Tabla 9. Perfil del cuarto encuestado. Fuente: elaboración propia*

**1ª Pregunta:** ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?

**Respuesta:** En conflictos latentes, muy locales, acciones antiterroristas, etc. Sobre todo, por el alto coste de estos dispositivos en el corto plazo.

**2ª Pregunta:** ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?

**Respuesta:** Los países menos desarrollados estarán más sometidos.

**3ª Pregunta:** ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?

**Respuesta:** Más desigualdad por un lado y menor coste en vidas humanas (para los países con estas tecnologías). Posible embrión de más conflictividad por parte de las sociedades sometidas.

**4ª Pregunta:** ¿Entraña riesgos el uso de estas tecnologías?

**Respuesta:** Sociedades con más poder, más injusticias cometerán.

**5ª Pregunta:** ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?

**Respuesta:** Dependerá del uso que se les dé, su uso puede ser honesto o totalmente reprobable.

## Quinto encuestado

<b>Formación / Cargo / Área de conocimiento</b>	General del Aire en la reserva
<b>Fecha</b>	04/01/2022

*Tabla 10. Perfil del quinto encuestado. Fuente: elaboración propia*

**1ª Pregunta:** ¿Cuál es el escenario futuro en el que serán empleadas las armas autónomas y los drones?

**Respuesta:** Actuaciones quirúrgicas antiterroristas, contra crimen organizado y contra líderes indeseables.

**2ª Pregunta:** ¿Cómo cree que afectaría el uso de estas tecnologías a la sociedad de un futuro?

**Respuesta:** Evitará despliegues sobre el terreno que son costosos y siempre con rechazo social.



**3<sup>a</sup> Pregunta:** ¿Afectaría de forma negativa o positiva el uso de estas tecnologías?

**Respuesta:** La cuestión no son las armas sino los principios y finalidad moral de quien las usa.

**4<sup>a</sup> Pregunta:** ¿Entraña riesgos el uso de estas tecnologías?

**Respuesta:** El riesgo está en que manos caigan

**5<sup>a</sup> Pregunta:** ¿Desde su punto de vista el uso de armas letales autónomas es moralmente correcto?

**Respuesta:** Tan moralmente correcto como la moralidad de quién decida usarlas, porque es imparable su futura existencia.