Revista de Bioética y Derecho
Perspectivas Bioéticas
www.bioeticayderecho.ub.edu - ISSN 1886-5887

DOSSIER SOBRE INTELIGENCIA ARTIFICIAL, ROBÓTICA E INTERNET DE LAS COSAS

Traducir el pensamiento en acción: Interfaces cerebro-máquina y el problema ético de la agencia

Translating thought into action: Brain-computer interfaces and the ethical problem of agency

Traduir el pensament en acció: Interfícies cervell-màquina i el problema ètic de l'agència

ANÍBAL MONASTERIO ASTOBIZA, TXETXU AUSÍN, MARIO TOBOSO, RICARDO MORTE FERRER, MANUEL APARICIO PAYÁ, DANIEL LÓPEZ *

* Aníbal Monasterio Astobiza. Investigador posdoctoral Gobierno Vasco en el ILCLI, UPV/EHU, Visitante académico en el Oxford-Uehiro Centre for Practical Ethics, University of Oxford, colaborador del IFS-CSIC. E-mail: anibalmastobiza@gmail.com.

* Txetxu Ausín. Grupo de Ética Aplicada GEA, Instituto de Filosofía, CSIC. E-mail: txetxu.ausin@cchs.csic.es.

* Mario Toboso. Departamento de Ciencia, Tecnología y Sociedad, Instituto de Filosofía, CSIC. E-mail: mario.toboso@csic.es.

* Ricardo Morte Ferrer. Vocal LI²FE (Laboratorio de Investigación e Intervención Filosófica y Ética). E-mail: ricardo63@autistici.org.

* Manuel Aparicio Payá. Departamento de Filosofía, Universidad de Murcia. E-mail: manuel.aparicio@um.es.

* Daniel López, Grupo de Ética Aplicada GEA, Instituto de Filosofía, CSIC. E-mail: daniel.lopez-castro@cchs.csic.es.

[1] Este trabajo se enmarca en los proyectos europeos INBOTS (780073) y EXTEND (779982), del Programa H2020. Aníbal Monasterio Astobiza agradece el patrocinio del Gobierno Vasco.

Resumen

En este artículo, nos proponemos dos objetivos: el primero, describir la teoría clásica de la agencia intencional y cómo la neurotecnología de las interfaces cerebro-máquina desafía los requisitos de la teoría clásica de la agencia y de la consciencia corporal. La neurotecnología de las interfaces cerebro-máquina funciona implantando electrodos directamente en el área de la corteza motora del cerebro que controla el movimiento, y está diseñada para detectar las señales neuronales asociadas con la intención de moverse, que son después decodificadas por un algoritmo en un computador en tiempo real. Así, una persona podría pensar en mover su pierna o su brazo y la máquina recibiría la información de su pensamiento para traducir el pensamiento en acción, mediante prótesis internas o exoesqueletos. Esto es posible y sus aplicaciones se proyectan tanto sobre la rehabilitación de la funcionalidad motora, como sobre la posibilidad de mejoramiento (*enhancement*) de las capacidades humanas. Ambas aplicaciones dan lugar a numerosas implicaciones éticas, pero destacamos principalmente una, que denominamos: el problema ético de la agencia. El segundo objetivo del artículo es explorar brevemente la ética algorítmica en el contexto de las interfaces cerebro-máquina y cómo se entienden en este ámbito la autonomía, la responsabilidad y la privacidad informacional. Finalmente, abogamos por la necesidad de un marco ético de principios que regule la neurotecnología, y en tal sentido apelamos a los nuevos neuroderechos.

Palabras clave: interfaces cerebro-máquina; neurotecnología; acción; pensamiento; intención; ética.

Abstract

The aim of this article is twofold: Firstly, we intend to describe the classical theory of intentional agency and to analyze how the neuro-technology of brain-machine interfaces (BCI) challenges the demands of that classical theory of agency and body consciousness. BCI neuro-technology works by implanting electrodes directly into the motor brain cortex that controls movement and detect neuronal signals associated with the intention to move, what is decoded by an algorithm on a computer in real time. Thus, someone could simply think about moving a leg or an arm and the tool (a prosthesis or exoskeleton) would receive the information to translate thought into action. This is yet feasible and its applications could involve rehabilitation of motor function and the possibility of enhancing human abilities. Both applications give rise to various several ethical implications but mainly to one that we call “the ethical problem of agency”. Secondly, we briefly explore the ethics of algorithms in the context of BCI neuro-technology and the way autonomy, responsibility, and informational privacy are understood. Finally, we advocate the need for an ethical framework of principles governing neuro-technology, such as the new neuro-rights.

Keywords: BCI; neurotechnology; action; thought; intention; ethics.

Resum

En aquest article, ens proposem dos objectius: el primer, descriure la teoria clàssica de l'agència intencional i com la neurotecnologia de les interfícies cervell-màquina desafia els requisits de la teoria clàssica de l'agència i de la consciència corporal. La neurotecnologia de les interfícies cervell-màquina funciona implantant elèctrodes directament en l'àrea de l'escorça motora del cervell que controla el moviment, i està dissenyada per a detectar els senyals neuronals associades amb la intenció de moure's, que són després decodificades per un algoritme en un computador en temps real. Així, una persona podria pensar a moure la seva cama o el seu braç i la màquina rebria la informació del seu pensament per a traduir el pensament en acció, mitjançant pròtesis internes o exoesquelets. Això és possible i les seves aplicacions es projecten tant sobre la rehabilitació de la funcionalitat motora, com sobre la possibilitat de millorament (*enhancement*) de les capacitats humanes. Totes dues aplicacions donen lloc a nombroses implicacions ètiques, però destaquem principalment una, que denominem: el problema ètic de l'agència. El segon objectiu de l'article és explorar breument l'ètica algorítmica en el context de les interfícies cervell-màquina i com s'entenen en aquest àmbit l'autonomia, la responsabilitat i la privacitat informacional. Finalment, advoquem per la necessitat d'un marc ètic de principis que reguli la neurotecnologia, i en tal sentit apel·lem als nous neuro-drets.

Paraules clau: interfícies cervell-màquina; neurotecnologia; acció; pensament; intenció; ètica.

1. Introducción: El ABC de las Interfaces cerebro-máquina

Las interfaces cerebro-máquina utilizan la información neuronal para el control de dispositivos externos como videojuegos, prótesis, cursores de computador, sillas de ruedas, domótica, exoesqueletos, armamento, etc. Habitualmente, se destaca la aplicación de las interfaces cerebro-máquina para tratar de mejorar la funcionalidad de personas con algún tipo de limitación del movimiento o condición neurodegenerativa como el Parkinson, si bien no son estas las únicas aplicaciones posibles

Hay distintos tipos de interfaces cerebro-maquina, como veremos en la sección *1.1.*, pero todo equipo y componentes de un interfaz cerebro-máquina, debe incluir:

- Sensores: para registrar la actividad eléctrica del cerebro
- Decodificador: un algoritmo que convierte la actividad eléctrica del cerebro en una señal de comando
- Efectores o actuadores: como un cursor de computadora, brazo robótico, prótesis…

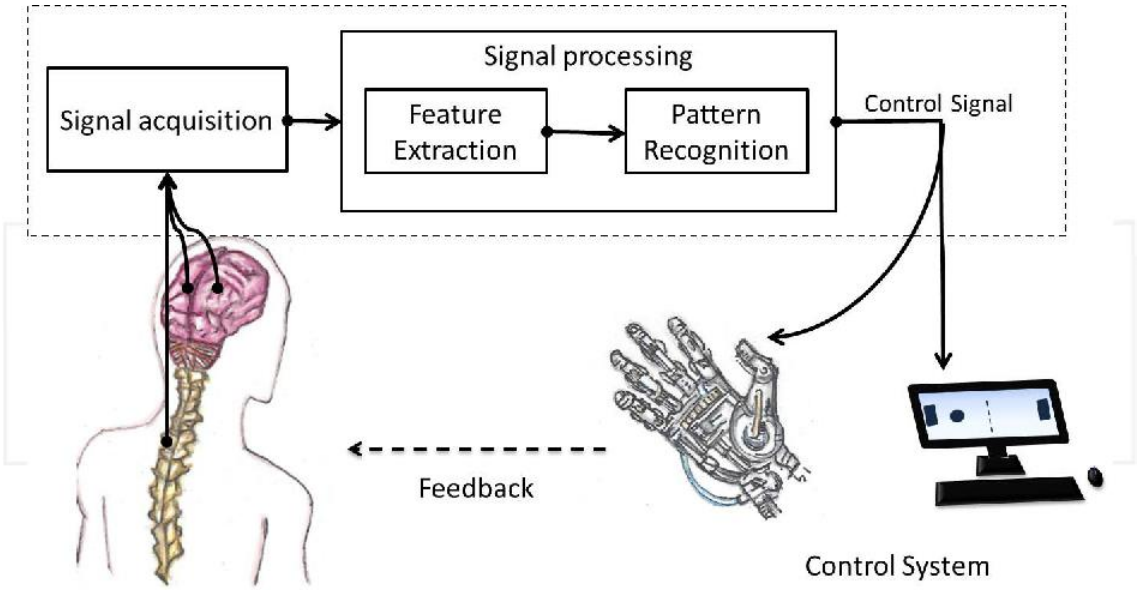

Figura. 1. Tomada de Ponce, Molina, Balderas y Grammatikou (2014).

Los sensores detectan la información (actividad eléctrica del cerebro) que puede ser actividad local de neuronas individuales o la actividad agregada de cientos, miles o millones de neuronas simultáneamente. El decodificador es un algoritmo matemático que procesa, filtra el ruido y amplifica la señal para operar el efector o actuador. La decodificación de la señal es clave para el funcionamiento efectivo del interfaz cerebro-máquina. Finalmente, el efector o actuador

viene a reflejar el tipo de aplicación de la interfaz cerebro-máquina. Interfaces cerebro-máquina se han aplicado en casos de limitaciones motoras, con el objetivo de facilitar la comunicación a través de la escritura en una pantalla de ordenador (Hochberg et al. 2006) o como respuesta física para controlar un efector (Wodlinger et al. 2015). Entre las interfaces más exitosas, porque se llevan décadas aplicando y comercializando, se encuentran las interfaces de estimulación, como por ejemplo, los implantes de cóclea (Gifford et al. 2008) en el caso de personas sordas; y la estimulación cerebral profunda para aliviar los síntomas en casos de enfermedad de Parkinson, distonia (Arle y Alterman 1999) o depresión (Trapp, Xiong y Conway 2018).

1.1 Taxonomía de las Interfaces cerebro-máquina

La mayoría de la neurotecnología en interfaces cerebro-máquina registra la actividad eléctrica del cerebro extracranealmente con electroencefalografía (EEG), magnetoencefalografía (MEG) o neuroimagen (IRMf etc.). La clasificación más básica que se puede hacer de las interfaces cerebro-maquina es en relación a su invasividad: existen interfaces cerebro-máquina invasivas y no-invasivas.

Esta distinción o clasificación depende sobre si la colocación o implantación de los electrodos requiere penetrar el sistema integumentario (piel).

A diferencia de las técnicas no-invasivas, los métodos invasivos tienen una mayor resolución espacial y especificidad temporal, pero tienen el riesgo de que es necesario implantar electrodos que penetren el sistema integumentario mediante operación quirúrgica. Los electrodos se pueden colocar también en el espacio subdural, entre el cerebro y los huesos del cráneo. En el caso más invasivo se puede implantar una micromatriz de múlti-unidades sobre la corteza como en el caso de la electrocorticografía (ECoG).

1.2. El futuro de las interfaces cerebro-máquina

Como hemos expuesto, la neurotecnología de las interfaces cerebro-máquina orienta uno de sus campos de aplicación a la rehabilitación de limitaciones funcionales, así como a aliviar los síntomas de condiciones neuropsiquiátricas como la depresión, a través de la estimulación. Además de estos campos de aplicación, la neurotecnología de las interfaces cerebro-máquina se aplica también fuera de un contexto clínico, y más allá de intervenciones rehabilitadoras puede

usarse para tratar de aumentar las capacidades cognitivas o el control sobre el entorno, como tecnologías para el mejoramiento humano (enhancement).

Desde la "neuroética" se advierte de algunos de los riesgos del uso de las neurotecnologías. Por ejemplo, las interfaces cerebro-máquina en su vertiente de estimulación pueden cambiar aspectos de la personalidad o del "yo" e incluso, como tienen como objetivo áreas especificas del cerebro, durante su uso pueden alterar los estados de ánimo, los deseos, la conducta e incluso los valores y la identidad personal (Schermer 2011). Es por todo ello, que es necesario realizar evaluaciones éticas sobre el acceso a las neurotecnologías, su utilización y sus consecuencias en el bienestar y en la autonomía de las personas.

La neurociencia detrás de las interfaces cerebro-máquina avanza tan rápido (Clausen 2009) que se pueden desarrollar aplicaciones como la comunicación de estímulos entre dos personas (Mashat, Li y Zhang 2017), la comunicación cerebro-a-cerebro a través de Internet (Rao et al. 2014), o la exploración del océano profundo y del espacio exterior con robots controlados por el pensamiento.

Investigadores del Instituto de investigaciones de Telecomunicaciones Avanzadas en Japón se han preguntado si una prótesis controlada por el pensamiento que trabaje junto con los brazos biológicos de una persona, puede dar a esa persona habilidades de multi-tarea superiores a las de una persona media (Penaloza y Nishio 2018).

Durante 20 sesiones los investigadores pedían a los participantes voluntarios completar tareas, a veces con sus brazos biológicos, y otras veces con el brazo robótico controlado con el pensamiento. En algunas sesiones las tareas las hacían simultáneamente, con sus brazos biológicos y el brazo robótico controlado con el pensamiento. Durante más de un 75% de las veces las dos tareas, una con los brazos biológicos y otra con el brazo robótico simultáneamente, se completaban con éxito. Esto sería imposible solo con los dos brazos biológicos.

Mejorar y aumentar las capacidades, no solo físicas, sino cognitivas puede ser una realidad con la neurotecnología. Así mismo, la posibilidad de *fusionar mentes* con la neurotecnología de interfaces cerebro-máquina podría considerarse también como una forma de aumentar o mejorar capacidades humanas o inclusive de otros animales no-humanos. La comunicación directa cerebro-a-cerebro humano es posible, pero también de un cerebro humano a un cerebro no-humano. Un grupo de investigadores (Yoo et al. 2013) ha conseguido trasmitir la actividad eléctrica del cerebro de un ser humano como "input" de entrada a la corteza motora de una rata durmiente que ha conseguido mover su cola. Se trata de una forma de comunicación entre dos especies diferentes mediada por las neurotecnologías.

Veamos ahora tres grandes áreas de potencial aplicación de la neurotecnología de las interfaces cerebro-máquina, cada una con sus oportunidades y riesgos: a) el entretenimiento, b) el campo militar y c) el consumo personal "háztelo tú mismo" (DIY).

a) El entretenimiento

En los últimos años el interés de aplicar la neurotecnología de las interfaces cerebro-máquina al entretenimiento (videojuegos) ha recibido un gran interés por la industria. De manera activa o pasiva la actividad del cerebro (i.e. P300 potencial relacionado a evento etc.) se pretende utilizar para propósito de control de avatares en juegos. El estado actual de la tecnología todavía es poco fiable, el hardware es limitado y la recepción de la señal es pobre. Pero a medida que se encuentren soluciones a estos obstáculos se podrá jugar con cascos de EEG controlando un avatar o personaje de videojuego con el pensamiento. La aplicación de las interfaces cerebro-máquina pueden revolucionar la forma en la que interactuamos con el entorno, y en concreto en su aplicación al entretenimiento las posibilidades están muy abiertas. Es posible controlar de manera directa un avatar inmerso en RV (realidad virtual), hacer de otro tipo de artes como el cine extremadamente interactivo o poder llegar algún día a jugarse deportes convencionales con sistemas neurrobóticos subrogados controlados por la mente. Los potenciales riesgos son serios y graves. Podemos llegar a "gamificar" la vida a tal extremo que nuestra actividad diaria sea un macro-espacio lúdico donde diferenciar la realidad online de la realidad offline sea cada vez más difícil.

b) El campo militar

La Agencia de Proyectos de Investigación Avanzados de Defensa de los EE.UU., mejor conocida por su acrónimo inglés DARPA y famosa por permitir desarrollar los primeros prototipos de Internet, tiene distintos programas de investigación para crear soldados con mejoras físicas y cognitivas.

El objetivo del programa NESD (Neural Engineering System Design: Véase, https://www.fbo.gov/index?s=opportunity&mode=form&id=8af4b29a3c98dfbb3479fd6d3177d7d0&tab=core&_cview=0) es desarrollar interfaces cerebro-máquina que permitan mejorar la comunicación de precisión entre el cerebro y el mundo digital. Cinco universidades han ganado la licitación pública y contratos con DARPA para crear métodos "wetware", interfaces cerebro-máquina que permitan convertir rápidamente la información eléctrica y química del cerebro en computación binaria para que sea procesada e interpretada por un ordenador. Por ejemplo, el grupo de la Universidad de Columbia trabaja en la corteza visual para crear un sistema bioeléctrico no-invasivo que permita que un ordenador vea lo que una persona ve. No solo es conveniente para el ejército hacer de los soldados más resilientes ante las graves consecuencias

de la experiencia extrema que supone entrar en el campo de batalla, mejorar la recuperación tras un desorden por estrés postraumático, lesión medular, déficits sensoriales, depresión, ansiedad… También es interesante para el futuro del ejército aprovecharse de las potencialidades que brinda la neurotecnología para entrenar, equipar y hasta incluso crear al mejor soldado.

c) El consumo personal (DIY)

Una de las aplicaciones de la neurotecnología de las interfaces cerebro-máquina que mayores riesgos entraña es el uso personal desde una cultura "hazlo tú mismo" (Do It Yourself). Movimientos como el "Quantified self" intentan utilizar la tecnología para obtener datos de la vida diaria de las personas, y cuantos más mejor, para de esta forma, aparentemente, tomar mejores decisiones relativas a la salud, el ocio, el trabajo etc. Pero esta cultura puede poner a las personas en riesgos injustificados (Wexler 2017). Apps para Smartphone que miden la presión arterial, el consumo de calorías… en definitiva, para optimizar la vida, son en cierta medida benignas, aunque impliquen una preocupación ética sobre la privacidad de los datos. Sin embargo, construir de manera amateur y casera neurotecnología para aplicártela o aplicársela a tus allegados puede suponer grandes peligros y riesgos (Snow 2015). La mínima seguridad desde la cultura DIY es uno de estos peligros. Aunque en la entrada del Oráculo de Delfos estaba grabada la inscripción "Conócete a ti mismo", este tipo de cuantificación del yo desde la tecnología no hace más felices o más eficientes a las personas, y además incluso pueden causar daño físico. Los biohackers juegan con moléculas intentando crear formas nuevas de vida o manipular su propia biología (Reagle 2019). La seña de identidad de la cultura DIY es la libertad y autonomía ilimitada. Libertad ilimitada para transformarse morfológicamente (Bostrom 2005), pero la exploración y la auto-experimentación que esta cultura tiene como filosofía rectora puede tener implicaciones éticas relevantes como, por ejemplo, daños personales y tragedias.

En la sección 2 trataremos otras implicaciones éticas relativas a la capacidad de lectura y predicción de las intenciones de los individuos para el control de la neurotecnología en casos de acciones delegadas: ¿cómo la predicción de la intención realizada por la interfaz cerebro-máquina se solapa con la intención genuina del agente? Conocer y entender cómo se produce técnicamente la implementación de la intención del agente mediada por una interfaz cerebro-máquina nos clarifica problemas como la tipificación de las acciones en el derecho, la responsabilidad de las acciones mediadas por interfaces cerebro-máquina etc.

En la sección 3 plantearemos un marco ético de regulación de las interfaces cerebro-maquina a la luz de los nuevos neuroderechos.

2. La teoría clásica de la agencia intencional

De acuerdo con la teoría clásica de las acciones intencionales (Searle 1983), se distinguen dos tipos de movimientos: sucesos (o meros accidentes) y acciones. Lo que distingue a ambos tipos son las intenciones. Las acciones tienen un antecedente causal que son los estados mentales intencionales y los sucesos o meros accidentes no. Un suceso o mero accidente es un empujón fortuito que recibo cuando estoy esperando en la cola del supermercado, mientras que una acción es un acto de conducta voluntario, esperar voluntariamente en la cola del supermercado porque deseo comprar un producto.

Los estados mentales intencionales tienen un rol causal en la generación de la conducta abierta. A su vez toda acción intencional o actuar intencionalmente, significa: tener un deseo (u otro estado mental intencional con similar rol causal en la generación de una acción), la creencia de que la acción nos conducirá a la satisfacción del deseo, razonamiento para combinar deseo y creencia dirigido hacia el objetivo de la acción, la habilidad para la acción, y, finalmente, la consciencia de la realización de la acción.

Así tenemos que principalmente la teoría clásica de la agencia diferencia de manera estricta entre acciones que tienen como antecedente causal a las intenciones (estados mentales) de meros sucesos o acontecimientos entre los que se pueden encontrar eventos como los reflejos, convulsiones, acciones durante el sueño, automatismos etc.

El problema con la teoría clásica de la agencia es que pone el énfasis en la historia causal de las acciones y no tanto en las capacidades de la persona en el momento de realizar la acción. A veces las personas no actúan, incluso aunque sus movimientos son causados por los estados mentales correctos, porque carecen de control sobre el movimiento. Es lo que se conoce en la literatura de investigación como "cadenas causales desviadas" (Mele 1987), argumento que critica los postulados de la teoría clásica de la agencia con respecto a lo que es una acción.

Imagina la siguiente situación hipotética que describe las "cadenas causales desviadas":

> *Pedro quiere y tiene la intención de matar a Juan, pero esto le pone nervioso y hace que su dedo tenga un tirón, mueva el gatillo de su pistola, el arma dispare y mate a Juan.*

La pregunta es, ¿realmente Pedro mató a Juan intencionalmente, con una acción intencional, o fue un mero accidente, una acción sin control intencional?

De acuerdo con la teoría clásica de la agencia intencional esto sería visto como una acción intencional. Está presente la intención de, en este caso, matar a Juan, pero en el momento de

hacerlo no fue una acción de la cual Pedro estuviera en total control, sino que parece que fue un espasmo muscular o reflejo. No había sentido pleno de la agencia.

2.1. Sentido de la agencia

Sentido de la agencia, en definitiva, propiedad del cuerpo (body ownership); es esencial para caracterizar las acciones intencionales.

El sentido de la agencia es la habilidad de una persona para controlar sus acciones (Haggard y Tsakiris 2009, 242). Una distinción relevante para entender el sentido de la agencia es "juicio de agencia" y "experiencia de agencia" (Synofzi ,Vosgerau y Newen 2008). El "juico de agencia" hace referencia a la capacidad que tenemos de establecer juicios conceptuales sobre si fuimos nosotros o no quienes realizamos una acción. La "experiencia de agencia" se refiere al juicio subjetivo y fenomenológico de estar realizando una acción y es eminentemente no-conceptual.

En condiciones normales la "experiencia de la agencia" es condición necesaria para que se pueda dar "juicio de agencia": mi creencia de haber tirado a canasta se subordina a mi sensación del tacto de la pelota de baloncesto. Estos dos factores crean nuestro sentido de la agencia, pero muchas veces en ciertas patologías o condiciones neuropsiquiátricas pueden estar disociadas. Pacientes con agnosonosia de su hemiplejia niegan que su extremidad afectada esté paralizada y creen que han realizado una acción cuando en realidad su extremidad está inmóvil (Piedimonte et al. 2016).

El avance y desarrollo de las interfaces cerebro-máquina y otras neurotecnologías emergentes no solo desafía los requisitos de la teoría clásica de la agencia, también de nuestro sentido de la agencia.

2.2. Acciones corporales básicas y acciones mediadas por interfaces cerebro-máquina

Las acciones corporales básicas, acciones reconocidas por la teoría clásica de la agencia, que intenten evitar las "cadenas causales desviadas" y den cuenta de nuestro sentido de la agencia; son puestas en tela de juicio por las acciones mediadas por interfaces cerebro-máquina. A esta nueva tipología de acciones las llamamos "acciones subrogadas".

Las acciones corporales básicas quedan más o menos recogidas y en cierta medida explicadas desde la teoría clásica de la agencia. Sin embargo, las acciones mediadas por interfaces cerebro-máquina, no.

¿Qué es una acción mediada por un interfaz cerebro-máquina? El padre de la teoría clásica de la agencia, Donald Davidson (1963), afirmó que toda acción es un movimiento corporal en última instancia. En el caso de una acción subrogada, una persona en estado de reposo puede producir cambios, eventos, en su entorno a través de una interfaz cerebro-máquina.

Las interfaces cerebro-máquina crean un nuevo tipo de acción que la teoría clásica de la agencia no da cuenta y esto supone un reto para las concepciones legales que sitúan al "movimiento corporal voluntario" como el criterio para asignar responsabilidades por las acciones, como veremos en la sección 3.1.

Una serie de preguntas fundamentales con consecuencias para la ética de la agencia y para poder entender esta nueva tipología de acción que se crea a través de la mediación de interfaces cerebro-máquina y que nosotros denominamos "acciones subrogadas", son: ¿Qué es realmente una intención?, ¿cómo la actividad eléctrica del cerebro se corresponde con una intención? y ¿cómo extraer de la señal correspondiente sus características?

Nosotros distinguimos entre las "acciones corporales básicas" de las que habla la teoría clásica de la agencia, y que hemos explicado brevemente en la sección 2, y las "acciones subrogadas". Este último tipo de acción se crea a partir de la mediación de interfaces cerebro-máquina para producir cambios/eventos en el mundo. Las "acciones subrogadas" tienen los mismos efectos físicos y resultados que las "acciones corporales básicas", pero a diferencia de estas últimas, no implican un movimiento corporal.

La filosofía de la agencia contemporánea ha distinguido, varias nociones de intención (Pacherie 2007):

- Intención P (presente): formadas en una situación específica y vinculadas a un objetivo específico
- Intención F (futura): formadas antes de la acción y vinculadas con procesos deliberativos
- Intención M (motora): monitoriza el movimiento corporal.

Atendiendo a las preguntas que hemos formulado más arriba -esenciales para poder entender la diferencias entre las acciones corporales básicas y las "acciones subrogadas"- tenemos que saber cómo se extrae la señal correspondiente de la actividad eléctrica del cerebro identificada con una intención, en otras palabras, qué es la codificación/decodificación neuronal.

La codificación equivale a una correspondencia entre el mundo externo y la actividad cerebral. Es, por lo tanto, de dirección mundo-a-mente. La decodificación, por su parte, es una correspondencia entre la actividad cerebral con el mundo externo. De dirección mente-a-mundo.

La decodificación puede ser pensada como "leer la mente", dado que las señales sobre la intención motora preceden al movimiento. La extracción de la señal para una interfaz cerebro-máquina es problemática. Y aquí está el problema que nosotros calificamos como: *el problema ético de la agencia*.

¿Cómo la intención genuina del usuario se adhiere con la señal que es procesada y predicha por el algoritmo de la interfaz cerebro-máquina?, ¿cómo la predicción de la intención de la interfaz cerebro-máquina se solapa con la intención del agente?

Por ejemplo, imagina leer en silencio, en tu habla interna, la siguiente frase:

"La casa roja del estanque azul"

Suponiendo que eres un hablante competente en castellano, y que has leído en tu habla interna esta frase, si con las técnicas disponibles de la neurociencia observáramos qué áreas del cerebro responsables del procesamiento del lenguaje se han activado en tu cerebro y las comparáramos con las áreas del cerebro responsables del procesamiento del lenguaje de otra persona activadas durante la lectura en su habla interna; es muy probable que las áreas activadas no sean las mismas dada la variabilidad individual en las estructuras cerebrales responsables del procesamiento del lenguaje.

La representación conceptual y lingüística que haga tu cerebro de "casa" no será similar a la representación que haga otra persona. Y lo mismo con los otros conceptos.

Esta variabilidad representacional de las intenciones se puede trasladar al campo motor.

Imagina leer en silencio, en tu habla interna, la siguiente frase:

"Revés de tenis"

Con esta frase leída en tu habla interna, ocurre algo parecido. La biomecánica del revés de tenis, de la supuesta personificación imaginada que has hecho de alguien haciendo un revés de tenis (o igual te has imaginado a ti mismo haciendo el revés), ha podido ser completamente distinta a la representación imaginaria de la biomecánica de un revés de tenis imaginado motóricamente por otra persona. Lo que queremos señalar es que la codificación y/o representación genuina de una intención motora o de otro tipo que va a ser procesada por el algoritmo de una interfaz cerebro-máquina es multifactorial con formato representacional altamente variable. En otras palabras, existe neurodiversidad.

Los actuales algoritmos detrás de las interfaces cerebro-máquina son constructos matemáticos que pueden emplear técnicas de aprendizaje automatizado por lo que no están pre-programados para tomar decisiones basadas en reglas secuenciales, sino que cambian sus decisiones de manera autónoma. El problema es que la resolución y especificidad de los algoritmos en el tratamiento de las señales cerebrales que procesan, filtran y amplifican, pueden incorporar funciones de auto-corrección o auto-completado que contaminen o por lo menos "enriquezcan" sin autorización la intención extraída de la actividad eléctrica del cerebro que no se corresponda con la genuina intención del usuario (Véase, Yuste et al. 2017).

Por otra parte, dado que la actividad eléctrica del cerebro se deja al albur de estos algoritmos, los cuales se aplican para extraer las intenciones de los usuarios, es muy probable que cuestiones de la ética de los algoritmos, como su opacidad, falta de transparencia etc.; se hayan de tener en cuenta para valorar la fiabilidad de las interfaces cerebro-máquina (Véase, Wolkenstein, Fox y Friedrich 2018). Que los algoritmos tengan acceso a los estados mentales de los usuarios puede tener implicaciones para la privacidad e incluso responsabilidad y autonomía de las personas. Porque si se pudiera dar una brecha de seguridad, *brainhacking* (o *brainjacking*), en el que un algoritmo fuera corrompido remotamente y repercuta en el control de un dispositivo , daría lugar a no saber cómo va funcionar el algoritmo. Podríamos pensar en escenarios donde se produce un daño por la acción de un dispositivo tecnológico y donde en realidad no ha sido controlado por el usuario de la interfaz cerebro-maquina, por el cerebro del usuario en definitiva, sino porque alguien ha manipulado el algoritmo que, en última instancia, controla el dispositivo tecnológico. La cuestión de la responsabilidad de las personas a la luz de las interfaces cerebro-maquina adquiere una nueva dimensión. Esto es *el problema ético de la agencia*.

3. Ética de las interfaces cerebro-máquina

Para evitar *el problema ético de la agencia* y otros (i.e. *brainhacking* o *brainjacking*) es necesario tener en cuenta la ética de los algoritmos aplicada al contexto de las interfaces cerebro-máquina, pero también la diferencia cualitativa entre las acciones básicas corporales y las "acciones subrogadas", mediadas por las interfaces cerebro-máquina.

La ética de las interfaces cerebro-máquina se puede ver como un campo de aplicación específico de la ética de datos que se subdivide en tres grandes áreas de acuerdo con Floridi y Tadeo (2016):

1. La ética de datos (cómo los datos se adquieren, tratan y almacenan)

2. La ética algorítmica (cómo la IA, aprendizaje máquina y robots interpretan los datos)
3. La ética de las prácticas (desarrollar códigos de buenas prácticas para profesionales en esta nueva ciencia de datos)

La ética algorítmica en el contexto de las interfaces cerebro-máquina es un tema poco explorado. En otro lugar, uno de nosotros (Monasterio Astobiza 2017) ha hablado sobre la ética algorítmica en general. Aquí hablaremos de cómo la ética algorítmica se incardina dentro del contexto de las interfaces cerebro-máquina porque puede dar lugar al *problema ético de la agencia* que describíamos más arriba: el problema de cómo establecer fidedignamente la correspondencia entre la predicción de la intención que hace la interfaz cerebro-máquina con la intención genuina del agente.

Un algoritmo es una lista de instrucciones. Una secuencia de pasos para hacer algo (Hil 2015, 39). Existen muchas clases de algoritmos y distintas aplicaciones para diferentes campos y tareas, pero son tres las características o propiedades definitorias:

a. Universalidad
b. Opacidad
c. Impacto

Son universales los algoritmos porque están detrás de la operatividad de múltiples sistemas tecno-sociales que nos influyen en nuestro día a día. Por ejemplo, los algoritmos están detrás del control de las señales de tráfico, sistemas de recomendación de música y películas etc. Son opacos porque parece que hablan un lenguaje arcano, ininteligible para la mayoría de los mortales[1]. Los algoritmos son complejos incluso para los matemáticos e ingenieros que los diseñaron. Y finalmente, tienen un impacto en la vida de las personas y como dice O´Neil (2016) convertirse en una amenaza para nuestra democracia y sociedad.

Si trasladamos estas características de los algoritmos en general a los algoritmos que se aplican en las interfaces cerebro-máquina, tenemos *el problema de la agencia* que es, en última instancia, cómo los algoritmos procesan los datos (actividad eléctrica del cerebro) para el control de un dispositivo tecnológico. Si, en resumidas cuentas, una interfaz cerebro-máquina obtiene la actividad eléctrica generada por el usuario, la decodifica y procesa (por los algoritmos), y la dirige a un ordenador para el control y comando de un dispositivo tecnológico; nos encontramos que los

[1] Aunque como bien nos ha señalado un revisor anónimo esta afirmación debe ser matizada. Es verdad que no todos los algoritmos son opacos en el sentido de ser compleja su interpretación; véase, Lipton 2016. Y cuando se usa el término "opaco" se ha de dejar claro que también se puede hacer referencia a secretos industriales (Laat 2018).

algoritmos entran en juego en distintas fases de este proceso: en la extracción y clasificación, y en el control del dispositivo tecnológico.

En cada una de estas fases existen riesgos de seguridad y una de las tres características que definen a todo algoritmo viene a ser crucial: la opacidad. La falta de transparencia puede llevarnos no solo al *problema ético de la agencia*, sino también a la privacidad informacional, a la corrupción y contaminación de la señal que finalmente tendrá el control del dispositivo tecnológico y otros problemas como la autonomía de las acciones.

Ahora, veamos cómo las "acciones subrogadas", acciones mediadas por una interfaz cerebro-máquina, se tipifican en el derecho y la necesidad de unos nuevos neuroderechos humanos.

3.1. Las acciones en el derecho

Como decíamos más arriba, las "acciones subrogadas" no solo desafían la teoría clásica de la agencia intencional, pueden dar lugar también a lo que hemos llamado *el problema ético de la agencia*: cómo establecer fidedignamente la correspondencia entre la predicción de la intención que hace la interfaz cerebro-máquina con la intención genuina del agente. Pero es que resulta, que las "acciones subrogadas" también chocan con la perspectiva tradicional del derecho.

Cuando se habla de derecho hay que recordar que no es un cuerpo monolítico que existe y aplica por igual en distintas parte del mundo. No hay un sistema legal universal y existen diversas jurisprudencias. Pero muchos códigos y sistemas legales, tipifican o prohíben actos. Y por actos el derecho en general entiende, "movimientos corporales voluntarios".

El problema es que la mediación de las interfaces cerebro-máquina en las "acciones subrogadas", y la ausencia en éstas de movimiento corporal voluntario, dificultan la adscripción de responsabilidad. La doctrina jurídica no alcanza a subsumir estas acciones que, dadas las implicaciones éticas de los algoritmos, ponen en tela de juicio la autonomía de las personas y la privacidad informacional. Para muchos autores es muy posible que las "acciones subrogadas" no sean acciones en un sentido legal estricto, porque para adscribir responsabilidad es necesario identificar un movimiento corporal. Para otros autores, el dispositivo puede entenderse como parte del cuerpo. De lo que no cabe duda es que el potencial desarrollo de la neurotecnología y la consideración de las "acciones subrogadas", coloca a la tipificación ordinaria del derecho sobre la naturaleza de la acción en una situación de incertidumbre.

3.2. Neuroderechos humanos

Nuestra capacidad de medir, registrar, monitorizar y manipular la actividad del cerebro es a día de hoy excepcional. Este hecho ha dado lugar a la necesidad de considerar un nuevo marco de derechos humanos sobre el que hacer frente a todos estos retos como por ejemplo la seguridad, integridad física y mental, privacidad informacional (cognitiva) etc. (Casabona 2013).

Marcelo Ienca y Roberto Andorno (2017) defienden la creación de nuevos derechos humanos: los neuroderechos. En este entorno de avance rápido de la neurotecnología el último bastión de la libertad humana, la mente o consciencia, está en peligro. La neurotecnología puede acceder a nuestros estados mentales "leerlos" e incluso manipularlos.

En dicha propuesta de neuroderechos se incluyen:

- *Libertad cognitiva*: decisiones libres y competentes en el uso de interfaces cerebro-máquina y derecho a que el estado, corporaciones o actores maliciosos no manipulen los estados mentales.
- *Privacidad mental*: derecho a proteger a las personas del acceso no autorizado a sus datos cerebrales
- *Integridad mental*: el derecho a la integridad mental reconocido por el derecho internacional como la promoción de la salud mental debe expandirse e incluir el derecho a la no manipulación de la actividad mental por la neurotecnología.
- Continuidad de la identidad personal y vida mental: derecho a no alterar la continuidad de la identidad personal y vida mental por terceras partes.

Conclusiones

En este artículo, nos hemos propuesto dos objetivos: 1) describir brevemente la teoría clásica de la agencia intencional y analizar cómo la neurotecnología de las interfaces cerebro-máquina desafía los requisitos de la teoría clásica de la agencia y consciencia corporal. Esto da lugar a diversas implicaciones éticas relacionadas con la agencia, pero principalmente a una que hemos denominado: *el problema ético de la agencia*. 2) explorar brevemente la ética algorítmica en el contexto de las interfaces cerebro-máquina y cómo se entienden la autonomía, la responsabilidad y la privacidad informacional. Finalmente, vemos necesario la ampliación del marco de derechos humanos para incluir a los neuroderechos y seguir protegiendo las libertades y derechos fundamentales ante el avance y desarrollo de la neurotecnología que supone un claro riesgo a

nivel individual y un reto importante para la sociedad. Por todo ello, y para evitar lo que hemos denominado como el problema ético de la agencia, creemos que es necesario realizar un mayor esfuerzo teórico en el análisis de las implicaciones neuroéticas de las tecnologías emergentes que hacen uso de algoritmos y diseñar estrategias que promuevan sistemas centrados en el ser humano, comprensibles, transparentes, predecibles y controlables. Solo de esta manera se podrá incrementar la seguridad y conseguir innovaciones más efectivas.

Referencias

- ARLE J. y ALTERMAN R. 1999, "Surgical options in Parkinson's disease". *Med. Clin. North. Am*. 83, pp. 483–98.
- BOSTROM N. 2005, "In Defense of Posthuman Dignity". *Bioethics*. 19, 3, pp. 202–214.
- CASABONA C. M. 2013, "Consideraciones jurídicas sobre los procedimientos experimentales de mejora (*enhancement*) en neurociências". *Percurso Acadêmico*, Belo Horizonte, v. 3, n. 5, pp 80-107.
- CLAUSEN J. 2009, "Man, machine and in between". *Nature*. 457, (7233): pp.1080–1.
- DAVIDSON D. 1963, "Actions, reasons, and causes". *Journal of Philosophy* 60 (23), pp. 685-700.
- FLORIDI L. y TADDEO M. 2016, "What is data ethics?" *Phil. Trans. R. Soc. A*, vol. 374, no. 2083 20160360.
- GIFFORD R. et al. 2008, "Speech recognition materials and ceiling effects: considerations for cochlear implant programs". *Audiol Neurootol*. 13(3),pp. 193-205.
- HAGGARD P. y TSAKIRIS M. 2009, "The Experience of Agency Feelings, Judgments, and Responsibility". *Current Directions in Psychological Science*, 18(4), pp. 242–246.
- HIL R. 2016, "What an algorithm is?" *Philosophy and Technology* 29, 1, pp. 35-59.
- HOCHBERG. L. et al. 2006, "Neuronal ensemble control of prosthetic devices by a human with tetraplegia". *Nature*, 442, 7099, pp. 164–71.
- IENCA M. y ANDORNO R. 2017, "Towards new human rights in the age of neuroscience and neurotechnology". *Life Sci Soc Policy*. 13, (1) 5.
- LAAT P. 2018, "Algorithmic decision-making based on machine learning from Big Data: Can transparency restore accountability?" *Philosophy & Technology* 31, 4, pp 525–541.

- LIPTON Z. 2018 "The mythos of model interpretability" Queue - Machine Learning 16 3 pp 1-27.
- MASHAT M. , LI G. y ZHANG D. 2017, "Human-to-human closed-loop control based on brain-to-brain interface and muscle-to-muscle interface". *Scientific Reports*. 7, 11001.
- MELE A. 1987, "Intentional Action and Wayward Causal Chains: The Problem of Tertiary Waywardness". *Philosophical Studies*. 51, 1,pp. 55-60.
- MONASTERIO ASTOBIZA A. 2017, "Ética algorítmica: Implicaciones éticas de una sociedad cada vez más gobernada por algoritmos". *Dilemata*, 24, pp. 185-217.
- O´NEIL C. 2016, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. New York. Crown Publishing Group.
- PACHERIE E. 2007, "The sense of control and the sense of agency". *Psyche*, 13(1), pp. 1–30.
- PENALOZA C. y NISHIO S. 2018, "BMI control of a third arm for multitasking". *Science Robotics*. 3, 20, eaat1228.
- PIEDIMONTE A. et al. 2016, "From intention to perception: The case of anosognosia for hemiplegia". *Neuropsychologia*. 1, 87, pp. 43-53.
- PONCE P., MOLINA A., BALDERAS D. y GRAMMATIKOU D. 2014, "Brain Computer Interfaces for Cerebral Palsy" En *Cerebral Palsy - Challenges for the Future* (eds) Emira Svraka InTech.
- RAO R. et al. 2014, "A Direct Brain-to-Brain Interface in Humans". *PLOS One*. 5, https://doi.org/10.1371/journal.pone.0111332.
- REAGLE J. (2019), *Hacking Life: Systematized Living and Its Discontents.* Cam. Mass. MIT Press.
- SEARLE J. 1983, *Intentionality: An Essay in the Philosophy of Mind*. Cambridge: Cambridge University Press.
- SCHERMER M. 2011, "Ethical issues in deep brain stimulation". *Front Integr Neurosci*. 5: 17.
- SNOW J. 2015, "*Entering the Matrix: the Challenge of Regulating Radical Leveling Technologies*" Tesis de Máster, Monterey, CA: Naval Postgraduate School.
- SYNOFZIK M., VOSGERAU G. y NEWEN, A. 2008, "Beyond the comparator model: A multifactorial two-step account of agency". *Consciousness and Cognition*, 17, pp. 219–239.
- TRAPP N.T., XIONG W. y CONWAY C.R. 2018, "Neurostimulation Therapies". En: *Handbook of Experimental Pharmacology*. Springer, Berlin, Heidelberg.

- WEXLER A. 2017, "The Social Context of "Do-It-Yourself" Brain Stimulation: Neurohackers, Biohackers, and Lifehackers" *Front Hum Neurosci.* 10; 11:224.
- WODLINGER B. et al. 2015, "Ten-dimensional anthropomorphic arm control in a human brain–machine interface: difficulties, solutions, and limitations," *Journal of Neural Engineering*. 12, 1, 16011.
- WOLKENSTEIN A., Jox R. y Friedrich O. 2018, "Brain–Computer Interfaces Lessons to Be Learned from the Ethics of Algorithms". *Cambridge Quarterly of Healthcare Ethics*, 27, 4, pp. 635–646.
- YOO S. et al. 2013, "Non-Invasive Brain-to-Brain Interface (BBI): Establishing Functional Links between Two Brains". *PLOS One* https://doi.org/10.1371/journal.pone.0060410.
- YUSTE R. et al. 2017, "Four ethical priorities for neurotechnologies and AI". *Nature*. 551(7679), pp. 159–63.

Fecha de recepción: 20 de noviembre de 2018
Fecha de aceptación: 29 de enero de 2019