
Políticas conductuales inteligentes: Oportunidades y riesgos ético-políticos de la Inteligencia Artificial para el *nudging*

*Intelligent behavioral policies: Ethical and political opportunities
and risks of Artificial Intelligence for nudging*

LUCÍA ORTIZ DE ZÁRATE ALCARAZO

Universidad Autónoma de Madrid

Cómo citar/Citation

De Zárate-Alcarazo, L.O. (2023). La Políticas conductuales inteligentes: Oportunidades y riesgos ético-políticos de la Inteligencia Artificial para el *nudging*. *Revista Española de Ciencia Política*, 62, 67-93. Doi: <https://doi.org/10.21308/recp.62.03>

Resumen

El presente artículo explora las oportunidades y los riesgos de la aplicación de tecnologías de Inteligencia Artificial (IA) para la elaboración de políticas públicas conductuales basadas en el *nudging*. Este tipo de políticas, especialmente populares en países anglosajones, han demostrado ser eficaces para promover ciertas conductas y hábitos que se consideran positivos entre la ciudadanía. Sin embargo, no se encuentran exentas de controversia por sus posibles efectos nocivos en cuestiones fundamentales en los sistemas democráticos occidentales como la libertad individual y la autonomía humana. El artículo parte del marco para la implementación de *nudges* propuesto por el Behavioural Insights Team (BIT) y explora como la incorporación de la IA puede ofrecer oportunidades para hacer este tipo de políticas más eficientes y, por tanto, promover, por ejemplo, hábitos saludables entre la ciudadanía. Lo que denominaríamos políticas conductuales inteligentes. Al mismo tiempo, el texto aborda algunos de los problemas ético-políticos que surgen de la combinación de estas tecnologías con las políticas conductuales y que cuestionan su compatibilidad con los valores democráticos.

Palabras clave: Inteligencia Artificial, ética, políticas públicas, nudging, gobernanza algorítmica, políticas conductuales.

Abstract

The present article explores the opportunities and risks of applying Artificial Intelligence (AI) technologies to the development of behavioral public policy based on nudging. These types of policies, particularly popular in Anglo-Saxon countries, have proven to be effective

in promoting certain behaviors and habits considered positive among citizens. However, they are not exempt from controversy due to their potential harmful effects on fundamental aspects of Western democratic systems, such as individual freedom and human autonomy. The article is based on the framework for implementing nudges proposed by the Behavioral Insights Team (BIT) and explores how the incorporation of AI can offer opportunities to make these types of policies more efficient and, therefore, promote, for example, healthy habits among citizens. This is what we would call intelligent behavioral policies. At the same time, the text addresses some of the ethical and political issues that arise from the combination of these technologies with behavioral policies, questioning their compatibility with democratic systems.

Keywords: Artificial Intelligence, ethics, public policy, nudging, algorithmic governance, behavioral policy.

1. INTRODUCCIÓN

Las ciencias del comportamiento juegan cada vez un papel más relevante dentro de las áreas de la administración y las políticas públicas. Décadas de estudio en este campo han puesto de relieve que la racionalidad humana está lejos de ser perfecta, ilimitada o total (Pohl, 2022). Los seres humanos tendemos a presentar sesgos hacia el presente, sentimos aversión al riesgo, sufrimos «internalidades» (*internalities*) (Reimer y Houtmanfar, 2017), tendemos a dejarnos llevar por la inercia, etc. Estos y otros tipos de sesgos, heurísticas y «miserias» cognitivas funcionan en nuestra mente de tal modo que hacen imposible que nuestras decisiones sean perfectamente racionales y, por el contrario, nos sumergen en el escenario de la racionalidad limitada (Simon 1987, 1991, 1993, 1997, 2000). Las limitaciones cognitivas no son algo negativo *per se*, de hecho, son un rasgo adaptativo gracias al cual podemos desenvolvernos en nuestro día a día sin grandes dificultades. Sin embargo, al mismo tiempo, aceptar estas premisas ha puesto de relieve que los seres humanos con frecuencia no tomamos decisiones óptimas en aquellos asuntos que tienen que ver con nuestro propio bienestar, y mucho menos con el de la mayoría.

En este sentido, durante los últimos años, desde el ámbito de las ciencias sociales y las humanidades, se ha propuesto explorar de qué forma se podrían aprovechar estas limitaciones cognitivas para que, en vez de repercutir(nos) negativamente, estas estén orientadas hacia el bienestar individual y común. Actualmente, la manera más habitual de hacer uso de las ciencias del comportamiento en el sector público es a través del uso *nudges*. Por otro lado, de manera más o menos simultánea, el sector público se encuentra inmerso en otro tipo de transformación o revolución, «la revolución algorítmica». La «revolución algorítmica», algoritmización o gobernanza algorítmica (Just y Latzer, 2017) se refiere al fenómeno de transformación del sector público producido por la incorporación de nuevas tecnologías disruptivas, como la Inteligencia Artificial (IA) basadas en algoritmos inteligentes (Meijer y Grimmelikhuisen, 2020). Debido a la novedad de ambas cuestiones, tanto del *nudging*, como de la IA, ambos temas se

encuentran en fases de estudio relativamente tempranas. En este sentido, algunas de las cuestiones más discutidas al tratar estos temas tienen que ver con las implicaciones éticas de la introducción de ambos tipos de prácticas en el sector público, pero también con los retos asociados a su incorporación y adopción en la esfera pública.

Por ello, el objetivo de este trabajo es plantear de forma introductoria propuestas sobre el modo en que la IA podría ser aprovechada para la práctica del *nudging*, tanto en su implementación como en su uso (*intelligent nudging*), así como los problemas ético-políticos que pueden surgir de esta fusión. Con el fin de alcanzar este objetivo en la siguiente sección presentaremos el marco teórico del artículo, el cual incluye tanto la Teoría del *Nudging* y los tipos de *nudges* que se han identificado hasta el momento como una revisión de la literatura sobre *nudges* en el sector público. En la tercera sección se expone la metodología y en la cuarta los resultados del artículo donde plantearemos las distintas formas en las que la IA puede transformar el proceso de implementación y uso de *nudges*. Finalmente, las secciones 5 y 6 de este trabajo se centrarán en la discusión de las propuestas planteadas y en las conclusiones del estudio.

2. MARCO TEÓRICO

2.1. La Teoría del *Nudging*

Un *nudge* es una «intervención, por parte de una institución privada o pública, que afecta al comportamiento de las personas al mismo tiempo que preserva totalmente su libertad de elección»¹ (Sunstein, 2020a:4). El ejemplo canónico que se utiliza para explicar cómo funciona el *nudging* es el del GPS. Del mismo modo que el GPS sugiere la mejor ruta para llegar al destino que uno desea, sin que el hecho de rechazar la recomendación conlleve ningún tipo de consecuencia negativa; los *nudges* son un tipo de intervenciones que, aprovechando las limitaciones cognitivas humanas, nos ponen en el «buen camino» o la «buena dirección», sin consecuencias negativas para aquella persona que elige no seguir el camino recomendado. Por el carácter orientador de los *nudges* hacia ciertos objetivos o metas y, simultáneamente, su respeto por la autonomía y libertad individual de las personas, al *nudging* también se le denomina paternalismo liberal (*libertarian paternalism*) (Thaler y Sunstein, 2003; Sunstein y Thaler, 2003; Sunstein, 2014, 2020a). Por tanto, los *nudges* son mecanismos coercitivos alineados con los valores de las democracias liberales (Strabheim, 2020).

Como se ha señalado anteriormente, la práctica del *nudging* se basa en las premisas que aporta el campo de las ciencias del comportamiento, principalmente, el área de la psicología de la toma de decisiones. Y es que precisamente, con eso tiene

1. Traducido del original «A nudge is defined as an intervention, from either private or public institutions, that affects people's behavior while fully maintaining their freedom of choice» (Sunstein, 2020a:4).

que ver el *nudging*, con «dirigir» o, mejor dicho, «orientar», las decisiones de las personas en un sentido que les sea beneficioso. Así, el uso de los *nudges* surge de la aceptación de que las capacidades cognitivas humanas son «limitadas», y se podría entender como el intento por sacarle el máximo provecho a esta realidad (Gigerenzer, 2008, 2014). Este tipo de sesgos, heurística y atajos cognitivos, aunque algunos lo equiparan a poner el «modo automático» son el resultado de procesos evolutivos y adaptativos que nos permiten tomar decisiones rápidamente y que, con frecuencia, en contextos conocidos y repetitivos son de mucha utilidad. Esta sería la parte positiva de la racionalidad limitada. Sin embargo, estos mismos sesgos dificultan la toma de decisiones en contextos nuevos o donde la incertidumbre es relativamente alta, haciendo, en muchos casos, que las decisiones tomadas repercutan negativamente en nosotros mismos (Tversky y Kahneman, 2011; Thaler y Sunstein, 2008; Kahneman, 2011). Es, en este segundo escenario, dónde la aplicación de *nudges* resulta especialmente útil.

En función del tipo de estrategia que utilice o el tipo de heurística y/o limitaciones cognitivas en los que se basen, los *nudges* pueden tomar distintas formas. Concretamente se han identificado tres tipos de *nudges* (Güemes, 2017, 2021). En el primer tipo de *nudges* se encuentran todos aquellos que hacen referencia a conceptos, narrativas y maneras de entender el mundo que son fundamentales en el día a día de las personas. En este grupo, los *nudges* más usados son los que se encuentran enfocados hacia la resignificación o redefinición de ciertas prácticas. Algunos ejemplos serían el uso de porcentajes en vez de cifras concretas para mejorar la adhesión a cierto tipo de propuestas, por ejemplo, las de carácter fiscal; también encontraríamos en este grupo todas aquellas intervenciones orientadas a modificar patrones de conducta habituales en las personas, pero que por algún motivo se busca cambiar en un momento concreto. En este sentido se busca modificar tanto patrones aparentemente menores como hábitos más relevantes, es decir, desde dejar de fumar en espacios públicos hasta cambiar las prácticas habituales de festejo en grandes eventos.

En el segundo tipo de *nudges* encontraríamos los «que apelan a normas, valores o sentimientos de orgullo social donde se referencia claramente la conducta esperada y se utilizan aspectos de la psicología social que buscan que los individuos se ajusten al grupo» (Güemes, 2021:314). En este grupo de *nudges* se enmarcarían todas aquellas iniciativas orientadas, por ejemplo, a participar en las campañas de recaudación de impuestos, de reciclaje, el uso de mascarillas en espacio públicos, etc. Uno de los *nudges* más usados en este campo son los mensajes en los que se visibiliza, a través de imágenes o cifras, el alto nivel de adhesión ciudadana a ciertas prácticas. Finalmente, en el tercer tipo de *nudges* entrarían los que se basan de manera más directa y concreta en la inercia. En este grupo se engloban todas aquellas prácticas que aprovechan el automatismo que desarrollan los seres humanos en la realización de ciertas tareas, así como nuestra predisposición a elegir la opción más fácil y/o más accesible. Los *nudges* de este tipo más conocidos son, entre otros, la disposición de ciertos productos en los supermercados, el uso de determinados colores para mejorar las ventas de algunos artículos, la colocación de los alimentos más saludables en lugares más accesibles en los

comedores, la introducción de opciones seleccionadas por defecto, etc. (Sunstein, 2014, 2020b).

2.2. Nudges en el sector público

El uso de *nudges* es cada vez más frecuente en el sector público, una práctica que ha ido en aumento, principalmente, a partir de la década del 2010 y que se ha popularizado de manera más significativa en los últimos años. En esta sección presentamos una breve revisión de literatura sobre el uso *nudges* en la administración y las políticas públicas.

El auge de las ciencias del comportamiento ha convertido a este campo de estudio en un terreno fértil que permite aportar nuevas perspectivas a disciplinas afines como la ciencia política, las ciencias humanas y morales, etc. Este hecho ha producido un aumento de la producción académica en torno a este fenómeno, el *nudging* en el sector público, durante la última década. Por ello, la revisión de literatura que se ha realizado se ha limitado a todos aquellos artículos publicados entre el año 2010 y el 2021 incluidos en la base de datos de la Web of Science. Dado que el objeto de estudio de este trabajo son los *nudges* en el sector público, los artículos en los que estamos interesados son los publicados sobre el tema en revistas especializadas en las áreas de la administración pública, las políticas públicas y las relaciones internacionales. Otros criterios de búsqueda han sido la inclusión de la palabra «*nudge*», «*nudges*» y/o «*nudging*» en el abstract, las palabras clave y/o el título del artículo. Así pues, la sintaxis usada en la revisión de literatura ha sido la siguiente:

TS= ((«Nudges» OR «Nudge» OR «Nudging») AND («Public Policy» OR «Public Administration» OR «Political Science» OR «International Relations»)) OR TI=((«Nudges» OR «Nudge» OR «Nudging») AND («Public Policy» OR «Public Administration» OR «Political Science» OR «International Relations»)) OR AB=((«Nudges» OR «Nudge» OR «Nudging») AND («Public Policy» OR «Public Administration» OR «Political Science» OR «International Relations»)) OR AK=((«Nudges» OR «Nudge» OR «Nudging») AND («Public Policy» OR «Public Administration» OR «Political Science» OR «International Relations»)).

Después de eliminar todos aquellos artículos que aparecían duplicados y que no seguían los criterios de búsqueda mencionados anteriormente, la muestra final estaba compuesta por 41 documentos publicados en revistas de relaciones internacionales, administración y políticas públicas, en inglés (Tabla 1). En la Figura 1 se ilustra la distribución de las fechas de publicación de los artículos que componen la revisión de literatura. Podemos observar que la literatura académica sobre *nudges* empieza a ganar relevancia a partir del año 2015. Antes de esa fecha no se ha encontrado ningún artículo científico que cumpliera los criterios de búsqueda seleccionados. Al mismo tiempo, la Figura 1 ilustra el auge vivido en el estudio del *nudging* en el sector público a partir del año 2019, momento en el que se registra el máximo número de artículos publicados sobre el tema (15).

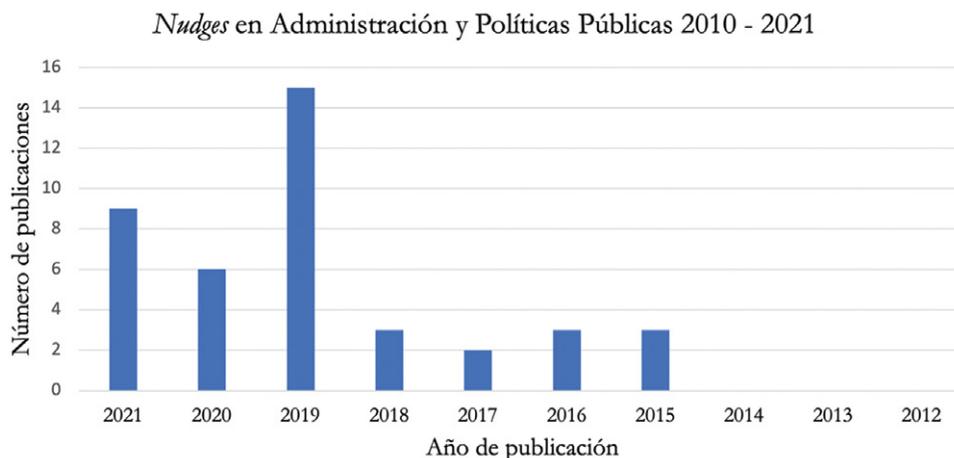
TABLA 1.
REVISIÓN DE LITERATURA SOBRE EL USO *NUDGES* EN EL SECTOR PÚBLICO.

Revista	Número de artículos científicos (2010-2021)
Policy and Politics	7
Public Administration Review	6
Policy Studies	5
Public Policy and Administration	3
Journal of European Public Policy	3
Public Management Review	2
Public Administration	1
Policy Studies Journal	1
International Review of Administrative Sciences	1
Policy and Society	1
Policy Sciences	1
International Review of Administrative Sciences	1
Administration and Society	1
Democratization	1
Evidence and Policy	1
Government Information Quarterly	1
American Review of Public Administration	1
Politics and Governance	1
American Political Science Review	1
International Public Management Journal	1
European Journal of Political Research	1
Total	41

Fuente: Elaboración propia.

Aunque la literatura sobre *nudges* en el sector público es relativamente variada, la mayor parte se centra en las oportunidades y los retos que puede suponer la incorporación de estas herramientas procedentes de las ciencias del comportamiento en el ámbito de la administración y las políticas públicas (Schmidt, 2017; Oliver, 2019; Tummers, 2019; Kuehnhanss, 2019; Strassheim, 2021). En estos artículos se suele introducir el concepto de *nudges*, su vinculación con el campo de las ciencias del comportamiento y las distintas áreas o políticas concretas que podrían verse beneficiadas por el uso de este tipo de herramientas conductuales. Por ejemplo, Einfeld (2019) relaciona el uso de *nudges* en el sector público con un aumento de las políticas públicas basadas en evidencias; Visintin *et al.* (2021) propone el uso de *nudges* para aumentar la eficiencia en el trabajo de los street-level bureaucrats en políticas de bienestar, etc.

FIGURA 1.

ARTÍCULOS CIENTÍFICOS SOBRE *NUDGES* EN ADMINISTRACIÓN Y POLÍTICAS PÚBLICAS, 2010-2021

Fuente: Elaboración propia.

Entre los campos donde más se ha estudiado el uso de *nudges* o, más bien, la posibilidad de introducir *nudges*, es en el área de la salud pública (Thomann, 2018; Esmark, 2019; Prainsack, 2020). En ellos se suelen ilustrar distintas formas en las que el *nudging* podría cambiar las prácticas relacionadas con la salud por parte de la ciudadanía (Vlaev *et al.*, 2016) de tal modo que estos pudiesen desarrollar fácilmente hábitos más saludables. También es interesante señalar que, aunque la búsqueda está limitada al año 2021, recientemente se ha observado un aumento en el número de artículos que promueven el uso de nudges para promover prácticas de salud pública acordes con las limitaciones y restricciones impuestas por los distintos gobiernos a raíz de la pandemia de la COVID-19 (Guan *et al.*, 2021).

También hay algunos artículos, aunque estos son minoritarios en los que se estudia la aprobación ciudadana del *nudging* a través de encuestas (Sunstein *et al.*, 2019), casos concretos en el uso de *nudges* (Gopalan y Pirog, 2017; Turner *et al.*, 2020) como, por ejemplo, en la donación de órganos (Moseley y Stoker, 2015); artículos en los que se sugieren cambios profundos en el sector público como resultado del *nudging* (Kasdan, 2019), etc. Finalmente, un número importante de artículos se centra en las limitaciones del *nudging* y en la necesidad de ir más allá de los enfoques tradicionalmente trabajados que suelen centrarse únicamente en el uso de *nudges* y olvidan otro tipo de herramientas conductuales (Mols *et al.*, 2015; Ewert, 2020; Ewert y Loer, 2021). Estos enfoques críticos también se encuentran en temas relacionados con la salud, por ejemplo, Prainsack (2020) alerta sobre los peligros de la personalización introducida por el *nudging* en las prácticas médicas por problemas relacionados con la violación de la privacidad de los pacientes y una posible disminución en la eficacia de los tratamientos médicos.

2.3. IA en el sector público

De forma más o menos paralela a la introducción del *nudging*, el sector público está viviendo en los últimos años una transformación que puede cambiarlo para siempre (de Sousa *et al.*, 2019). Esta transformación está relacionada con la incorporación de nuevos grupos de tecnologías disruptivas, como la IA, en el ámbito de la gobernanza (Margetts y Dorobantu, 2019), la administración y las políticas públicas (Valle-Cruz *et al.*, 2020; Meijer y Grimmelikhuijsen, 2020), y la consecuente algoritmización de las estructuras burocráticas, la creciente datafización (*datafication*) del sector público, etc. A partir de estos cambios se espera un incremento del potencial de acción de las instituciones públicas y el diseño de políticas públicas.

Debido a sus características propias, como el manejo de grandes cantidades de datos (*big data*), el uso de algoritmos inteligentes y el incremento significativo en las capacidades de procesamiento de datos (Russell y Norvig, 2016), la IA permite automatizar muchas tareas y operaciones, aumentando la eficacia y la eficiencia de las mismas, y al mismo tiempo, realizarlas con un grado de personalización muy alto, como, por ejemplo, sucede en el caso de los diagnósticos médicos (Sun y Medaglia, 2019) o en el seguimiento de ciertas enfermedades especialmente peligrosas (Greene-meier, 2014). Otras de las oportunidades más importantes que conlleva la introducción de IA en el sector público tiene que ver con las capacidades de este nuevo grupo de tecnologías para la detección en tiempo real de personas y/o hechos, así como la predicción de ciertos fenómenos. En este sentido son especialmente conocidas las aplicaciones de la IA al ámbito de la seguridad y la justicia (Gasser y Almeida, 2017; Misuraca y Van Noordt, 2020), donde la IA permitiría, por ejemplo, predecir las probabilidades de reincidencia de algunos delincuentes, detectar lugares especialmente peligrosos o hechos delictivos en tiempo real (Quest *et al.*, 2018).

Sin embargo, todos estos cambios y transformaciones que prometen mejorar el sector público también conllevan numerosos problemas que dificultan seriamente la materialización de los planes de algoritmización. Más allá de los retos organizativos y administrativos implícitos en cualquier proceso de modernización del sector público, el tránsito hacia un nuevo modelo de gobernanza como el de la gobernanza algorítmica implica una serie de problemas ético-políticos que no pueden pasarse por alto si aspiramos a que este nuevo paradigma esté orientado en aras del bien común (Margetts y Dorobantu, 2019; Coeckelbergh, 2020). En este sentido, la resolución de ciertos desafíos éticos como la explicabilidad (De Zárate-Alcarazo, 2022), el respeto por la autonomía humana y la privacidad (Véliz, 2021), la justicia y la no discriminación (De Zárate-Alcarazo, 2023), la transparencia (Burrell, 2016), etc., es una condición indispensable para garantizar un uso justo de la IA en el sector público. Además de estos, otro problema ético-político que ponen directamente en cuestión la viabilidad de la gobernanza algorítmica es la compatibilidad real de la transición tecnológica con la crisis climática (Almazán, 2021). Estas cuestiones, entre otras, ponen directamente en riesgo la viabilidad del uso masivo de la IA en el sector público.

En vistas a la resolución de estos problemas ético-políticos la Unión Europea (UE), lleva trabajando desde el año 2018 en un plan coordinado para la IA europea (Comisión Europea, 2018). El objetivo es garantizar que el uso de la IA en suelo europeo esté regido por un marco ético y legislativo común que permita asegurar los derechos de la ciudadanía europea en toda la Unión. Así, en el año 2019 la Comisión Europea publicó el Libro Blanco de la IA en la UE (Comisión Europea, 2019) donde recogía los principios éticos que debe cumplir cualquier sistema de IA que sea aplicado dentro de nuestras fronteras. Estos principios son: la explicabilidad de la IA (XAI), el no-daño, el respeto por la autonomía humana y la justicia. Para poder garantizar el cumplimiento de estos principios la UE ha desarrollado la primera legislación a nivel mundial en materia de IA (Comisión Europea, 2021). Dicha legislación diferencia entre cuatro tipos de IA en función del nivel de riesgo (inaceptable, alto, medio y bajo) que conlleva su uso en relación con la posible violación de derechos y libertades fundamentales. En base a la presente legislación, cada tipo de IA está sometido a unos controles regulatorios más o menos estrictos.

Todas estas características propias de la IA (eficacia, eficiencia, personalización, automatización, detección, predicción, etc.), así como el contexto que la hacen posible la convierten en un buen complemento para impulsar el uso de *nudges* en el sector público. Sin embargo, la presencia de distintos problemas éticos, como los mencionados anteriormente, supone un importante obstáculo de cara a hacer que las políticas conductuales inteligentes sean realmente viables dentro del marco de derechos y libertades que caracterizan a las democracias liberales europeas. Esta cuestión se desarrollará con más detalle en la sección 5 dedicada a la discusión de los resultados.

3. METODOLOGÍA

Cuando hablamos de usar *nudges* en el sector público, y también en el sector privado, el marco más usado es el desarrollado por el Behavioural Insights Team (BIT) (2014), que se compone de cuatro pasos. Normalmente se conoce con el nombre de EAST en referencia al método de intervención que implica y que es, a su vez, el paso central del marco del *nudging*. El término EAST hace referencia a los términos: fácil, atractivo, social y preciso (en inglés, *easy, attractive, social* y *timely*). Aunque también se han propuesto algunas versiones alternativas o pequeñas modificaciones como el FEAST, donde a la tradicional formula EAST se le añade el elemento de la diversión (*fun*) (Sunstein, 2020a). En este trabajo vamos a trabajar con el marco desarrollado en 2014 por el BIT, por ser la versión más extendida en lo que se refiere a la implementación de *nudges*.

A su vez, la incorporación de la IA al sector público puede suponer una oportunidad sin precedentes para mejorar la práctica del *nudging*, por la que cada vez más gobiernos apuestan (Sousa Lourenço *et al.*, 2016). Al igual que sucede en el caso del ciclo de políticas públicas (Valle-Cruz *et al.*, 2020), el gobierno (Margetts y Dorobantu, 2019) y las administraciones públicas (Valle-Cruz *et al.*, 2020)), la IA puede transformar el uso de *nudges* en el sector público. Las oportunidades asociadas a la IA como la automatización, la capacidad de detección y predicción, así como la personalización, pueden hacer que los *nudges*

aumenten su eficacia y eficiencia y dar lugar a lo que podría denominarse como *nudges* inteligentes. De esta forma estaríamos hablando de un nuevo tipo de intervenciones pertenecientes a un campo emergente como es el de las políticas conductuales inteligentes.

Así pues, las políticas conductuales inteligentes serían todas aquellas políticas que, basándose en las premisas y conocimientos procedentes de las ciencias del comportamiento, hacen a su vez uso de tecnologías basadas en algoritmos inteligentes, como es el caso de la IA (Guihot *et al.*, 2017). Para comprender como la IA puede transformar el ciclo de implementación de *nudges* primero es importante conocer cómo funciona dicho proceso. El marco para la implementación de *nudges* propuesto por el BIT se compone de cuatro pasos: definición de objetivos, comprensión del contexto, intervención y evaluación.

En lo que se refiere a la implementación y el uso de *nudges* creemos que el añadido de la IA puede transformar los cuatro pasos que componen el marco propuesto por el BIT para su implementación y, a su vez, los cuatro elementos que conforman el método EAST y que se corresponde con el tercer paso del ciclo, la intervención (Tabla 2). De este modo, la metodología seguida parte del marco teórico de aplicación del *nudge* propuesto por el BIT y analiza como la IA puede transformar cada uno de los cuatro pasos que componen el proceso de manera exploratoria. Este tipo de técnicas exploratorias son utilizadas con frecuencia en la investigación cualitativa sobre temas emergentes como son la IA y los *nudges* (Lune y Berg, 2017).

4. RESULTADOS

Políticas conductuales inteligentes

El primer paso en el proceso de implementación de intervenciones conductuales, como son los *nudges*, es el de la definición de objetivos. Este momento es similar al de otros procesos como puede ser el *agenda-setting* en el proceso de elaboración de políticas públicas (Harguindéguy, 2013) y en él lo que se busca es, principalmente, identificar los problemas más relevantes y los comportamientos asociados a dichas problemáticas. De este modo, es posible establecer qué tipo de cambio se busca producir en el comportamiento de las personas, en cuánto tiempo sería razonable observar dicho cambio, así como otras cuestiones similares (BIT, 2014). En este sentido, el añadido de la IA a la práctica del *nudging* podría resultar muy útil en este primer momento del ciclo de tal modo que el número de respuestas positiva a las políticas de estímulo creciese.

La aplicación de algoritmos inteligentes capaces de hacer seguimiento de los debates en redes sociales y detectar las emociones que generan en tiempo real, aunque puede conllevar serios riesgos relacionados con la violación de la privacidad de los ciudadanos, el condicionamiento de los debates públicos, etc., (Lazer, 2015). También puede ayudar a mejorar la eficacia, eficiencia y precisión de las decisiones que se toman en este paso del proceso de implementación de *nudges*, principalmente, mejorando la selección de los problemas y comportamientos a tratar. Además, la capacidad de predicción de muchos

Tabla 2.
NÚDGE S INTELIGENTES: OPORTUNIDADES Y RIESGOS

	Cuestiones relevantes		Oportunidades	Riesgos
Definición de objetivos	¿Qué problemas seleccionamos?	¿Qué comportamientos hay que cambiar?	Mayor: eficiencia, precisión, rapidez en la detección de problemas y comportamientos. Añadido: personalización y predicción de los efectos del nudging.	Problemas de privacidad de los datos, disminución de las libertades de la ciudadanía, determinismo, sesgos y sobre dimensionado de ciertos colectivos.
Comprensión del contexto	¿Cómo podemos conocer el entorno?	¿Cómo identificamos elementos desencadenantes de conductas inapropiadas en el entorno?	Mayor: rapidez, eficiencia, eficacia, precisión en la identificación del entorno. Añadido: abaratamiento del proceso de identificación al evitar desplazamientos.	Problemas de privacidad de los datos, disminución de las libertades de la ciudadanía, sesgos, opacidad, deshumanización y robotización del trabajo.
Intervención	¿Cómo podemos hacer esta tarea/servicio más fácil?	¿Cómo podemos hacer esta tarea/servicio más atractiva?	Mayor: detección de dificultades en la realización de una tarea, evaluación y predicción de soluciones. Añadido: capacidades tecnológicas habilitadoras. Mejor: evaluación de mensajes atractivos, detección de emociones. Añadido: personalización del contenido del mensaje, capacidades tecnológicas habilitadoras. Mejor: conocimiento de las estadísticas y los datos, personalización del mensaje, eficacia, precisión. Añadida: capacidades tecnológicas para habilitar nuevos espacios de socialización.	Problemas de privacidad, recolección de datos, brecha digital, determinismo, disminución de la confianza ciudadana, problemas éticos y legales, sesgos.
Evaluación	¿Cómo podemos hacer esta tarea/servicio más precisa?	¿Cómo sabemos si la aplicación del nudge a funcionado?	Mayor: personalización, detección del momento adecuado, seguimiento de las personas, detección de emociones. Mayor: eficacia en las métricas, identificación de problemas. Añadido: evaluación en tiempo real de los efectos del nudging.	Opacidad algorítmica, problemas de rendición de cuentas, alta dependencia tecnológica, sesgos.

Fuente: Elaboración propia.

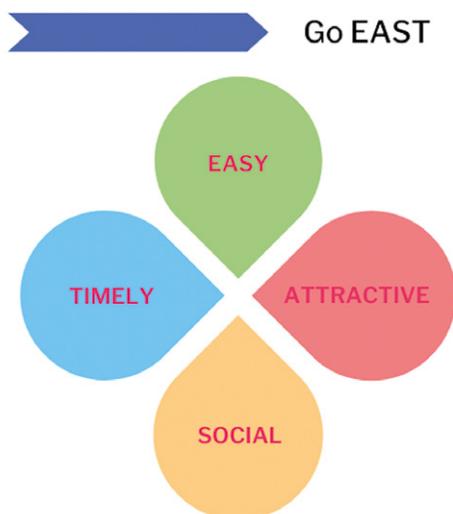
de los sistemas de IA también podría ayudar a transformar este paso del proceso, así como la totalidad del mismo, haciendo posible adelantar las posibles consecuencias de las intervenciones antes de que tengan lugar. De este modo, se podrían evitar posibles errores, malas elecciones, así como los costes asociados a todo ello.

El segundo momento en el proceso del *nudging* se encuentra enfocado en la comprensión del contexto en el cual se produce el comportamiento y/o hábito que se busca orientar ya que, con frecuencia, las prácticas humanas son altamente dependientes del contexto y muchos *nudges* hacen uso de la disposición y colocación de elementos en un espacio determinado. Inicialmente, la forma habitual de conocer el contexto ha sido la de enviar personal a los escenarios seleccionados para poder estudiar el entorno en detalle y conocer las particularidades de las personas afectadas. En este sentido, la introducción de la IA y el tránsito hacia la aplicación de políticas conductuales inteligentes podría simplificar, agilizar y mejorar este paso de forma significativa a través del uso de *softwares* de reconocimiento facial, de objetos y del entorno, tecnologías de procesamiento del lenguaje natural y otros sistemas similares a los que ya se utilizan, por ejemplo, para realizar diagnósticos médicos (Sun y Medaglia, 2019). La aplicación de este tipo de tecnologías podría ser muy útil para identificar y evaluar el contexto en el cual se suele producir el comportamiento a orientar sin necesidad de que se desplacen personas. De este modo, además de las ventajas anteriormente mencionadas, se podrían abaratar significativamente los costes asociados a este paso.

Para ilustrar las oportunidades de la IA en este segundo momento del proceso de implementación de *nudges* puede ser útil referirse a uno de los ejemplos más usados en este campo, como es el del pago de impuestos. Recientemente, se ha demostrado que el uso de *nudges* puede ser útil para aumentar el número de personas que pagan impuestos a tiempo y siguiendo los cauces administrativos y legales más habituales (Hallsworth *et al.*, 2017). En este sentido, la incorporación de la IA al *nudging* podría usarse, por ejemplo, para detectar qué personas suelen eludir el pago de impuestos, identificar posibles patrones en sus perfiles, mejorar la comprensión de los motivos por los cuales deciden no pagar impuestos (a través del análisis del texto y emociones), cruzar datos de naturaleza emocional con datos fiscales, laborales o de otro tipo, etc., de tal modo que la comprensión del contexto fuera mucho más precisa.

El tercer paso del proceso es la intervención, y es el que propiamente constituye la aplicación del *nudging* y, por tanto, el núcleo central del marco desarrollado por el BIT. Como se ha mencionado anteriormente, este paso se conoce como EAST y consta de cuatro elementos a tener en cuenta y de cuya aplicación depende, *a priori*, la eficacia del *nudge* (Figura 2). A su vez, cada uno de los elementos del momento de la intervención, EAST, se basa en distintos sesgos cognitivos y heurísticas humanas. Esta característica es lo que convierte a los *nudges* en herramientas cognitivas que permiten orientar las prácticas humanas y transformar los hábitos de la ciudadanía en sentidos que les son más beneficiosos, al mismo tiempo que respetan su derechos y libertades (Behavioural Insights Team, 2012, 2014; Sunstein, 2020a). En la aplicación de cada uno de estos cuatro elementos la incorporación de la IA también puede ser útil de tal forma que mejorase el paso de la intervención en su totalidad.

FIGURA 2.
MÉTODO EAST PARA LA IMPLEMENTACIÓN DE *NUDGES*



Fuente: Elaboración propia.

A continuación, haremos un repaso por cada uno de estos elementos y como la IA puede ayudar a transformarlos a mejor. La primera premisa del *nudging* es que la intervención tiene que ser fácil. Uno de los principales descubrimientos procedentes de las ciencias del comportamiento es que, con frecuencia, las personas dejamos de hacer ciertas tareas debido a pequeñas cuestiones que, por algún motivo, nos resultan tediosas (Sunstein, 2020b). En la mayor parte de estos casos las implicaciones no son muy importantes, pero puede haber situaciones en las que sí lo sea, de tal modo que facilitar ciertas tareas, o parte de ellas, puede marcar una gran diferencia y producir un cambio en los patrones de comportamiento. Algunos patrones que suelen repetirse y que responden a este tipo de lógica son, por ejemplo, la dificultad para cancelar la suscripción al gimnasio cuando esta se renueva automáticamente de forma mensual, mantener los ajustes que vienen por defecto en los teléfonos móviles cuando los compramos o no cambiar las tarifas de la energía, la luz o el teléfono. Estos ejemplos que pueden parecer menores, también se reproducen a mayor escala, por ejemplo, en el caso de la donación de órganos o la recepción de mensajes gubernamentales, etc. (Kahneman y Tversky, 1984; Madrian y Shea, 2001; Beshears *et al.*, 2009).

La forma más común de hacer más fácil una tarea o la adscripción a ciertas prácticas suele ser a través del uso de reglas por defecto, el uso de mensajes simplificados y la reducción de trabas (*friction costs*) (Behavioural Insights Team, 2014:9). Algunas evidencias procedentes del sector público demuestran la eficacia de este tipo de prácticas. Por ejemplo, en Alemania el hecho de registrar, por defecto, a parte de la ciudadanía en una tarifa basada en energía eléctrica hizo que el número de personas inscritas a lo largo del tiempo se multiplicase por diez. Esto significa que la mayor parte de estas

personas no se salieron del nuevo plan y decidieron quedarse con la opción marcada por defecto, que, a su vez, era la más sostenible (Ebeling y Sebastian, 2015). Por otro lado, en el año 2012 el gobierno del Reino Unido decidió empezar a inscribir automáticamente a sus trabajadores en un plan de pensiones (siempre con la opción de salirse del mismo libremente sin ningún tipo de consecuencia negativa). Esto ha tenido como resultado un aumento significativo del número de personas que pueden beneficiarse de estos servicios (Gobierno de Reino Unido, 2013).

En este sentido, la IA puede ser de gran utilidad por muchas razones. Primero, muchas de las tecnologías de IA pueden ser de gran ayuda a la hora de identificar cuáles son las principales dificultades que afrontan los ciudadanos en la realización de una tarea, algo que no siempre es fácil de determinar. Por ejemplo, el inscribirse en un plan de pensiones puede resultar difícil para algunas personas por la complejidad a la hora de entender sus implicaciones, la complejidad de los trámites, o simplemente por desconocimiento. Por ello, en lo que se refiere a mejorar la adhesión a ciertas prácticas, el uso de apps inteligentes y otras funciones derivadas de la mejorar en las capacidades tecnológicas puede ser de gran ayuda. Si podemos aceptar o cancelar nuestra suscripción a servicios públicos tan solo pulsando un botón, las posibilidades de rehuir ciertas prácticas por cuestiones de dificultad o tedio disminuyen (Bettinger *et al.*, 2012). La IA también ser de gran ayuda en la simplificación de los mensajes y aportando información útil por ejemplo a través del uso de *chatbots* o asistentes de voz (Di Prospero *et al.*, 2017).

Otro de los pilares del método EAST consiste en hacer que la adhesión a ciertas prácticas resulte atractiva para la ciudadanía. En este sentido el sector público se encuentra rezagado respecto del sector privado ya que este último es especialmente eficiente en la elaboración de grandes campañas de marketing. La tarea de hacer más atractivo un servicio o producto se suele realizar, principalmente, de dos formas. Primero, consiguiendo focalizar la atención en el producto o servicio que nos interesa promover. Esto se consigue haciendo visibles los costes y los beneficios de usar un producto o servicio a través de mensajes, colores, etc. Por ejemplo, a través del uso de etiquetas de color rojo en alimentos perjudiciales para la salud (Mehta y Zhu, 2009). En este sentido la IA puede ser especialmente útil para personalizar los mensajes de tal modo que cada persona reciba un mensaje adaptado a sus propias preferencias. La segunda forma de hacer más atractivo un servicio suele ser a través del diseño de mecanismos de incentivos y sanciones (estos últimos sin repercusiones legales), por ejemplo, transformando el sentido de algunos comportamientos a través de su «conversión» en juegos (*gamification*), inscribiendo a las personas que se registran para votar en las elecciones en una lotería, etc. Aquí la IA y el desarrollo tecnológico juegan un papel muy relevante ya que son elementos decisivos a la hora de habilitar la creación de plataformas interactivas que permitan que todos los ciudadanos participen en los «juegos», loterías, etc. Siguiendo con el ejemplo del pago de impuestos, una forma de incorporar la tecnología sería incluyendo a todas las personas que pagan impuestos, de forma automática, en una lotería.

Otro de los recursos cognitivos más eficaces y que con más frecuencia utilizamos es la validación social por parte del grupo al que pertenecemos. Por este motivo,

tendemos a reducir nuestro consumo energético o reciclar cuando sabemos que otras personas de nuestro entorno también lo hacen. Consumimos productos y servicios que sabemos que le han sido de utilidad a nuestros allegados y conocidos, u otros ciudadanos con los que nos sentimos en igualdad de condiciones, ya sea por pertenencia a una misma ciudad, provincia, género, clase social, etc. (BIT, 2014:28). En base a la influencia social, que muchas veces pasa desapercibida, algunos recursos útiles que son habituales en la práctica del *nudging* son: mensajes en los que se muestra el porcentaje de personas que realizan una práctica deseable, como reciclar o pagar impuestos (Hallsworth *et al.*, 2017), o la creación de redes de contacto personales a través de las cuales las personas pueden aportar sus impresiones sobre sus experiencia al hacer uso de un servicio o realizar una práctica concreta —algo que ha resultado muy útil en las prácticas relacionadas con la donación de órganos o en la asistencia a personas mayores. En este caso la incorporación de la IA puede ayudar a crear, facilitar y reforzar las redes de contactos a través del uso de algoritmos que recomienden personas que gustos y necesidades afines. También se pueden utilizar tecnologías de IA para proporcionar porcentajes detallados, y que hasta ahora eran difíciles de conocer, sobre la eficacia y la utilidad de ciertos servicios públicos, hacer recomendaciones mucho más personalizadas, etc.

Finalmente, para que los *nudges* sean eficientes es importante que estos sean precisos. Esto significa, principalmente, tres cosas. Primero, que se apliquen en el momento adecuado. Actualmente sabemos que las personas están más dispuestas a realizar cambios en sus formas de vida cuando viven momentos de transición como, por ejemplo, al mudarse de casa, casarse, tener un hijo, etc. (Thompson *et al.*, 2011). Por ejemplo, mandar un mensaje personalizado, o incluso estandarizado, 10 días antes de que venza el plazo para pagar una multa ha demostrado ser un método eficaz para aumentar el número de personas que pagan (BIT, 2012). El uso de tecnologías de IA en este punto podría facilitar, agilizar y ayudar a determinar con mayor precisión la situación anímica de los ciudadanos y, por tanto, su disposición a recibir cierta información.

Segundo, para mejorar la precisión del *nudging* es importante centrarse en los costes y beneficios inmediatos. Es sabido que las personas evaluamos de forma más acusada todo aquello que sucede en el presente y tendemos a minimizar las consecuencias futuras (Zauberman *et al.*, 2009). De ahí que tantas personas se decanten por estilos de vida poco saludables que implican el consumo de tabaco, grandes cantidades de alcohol y/o alimentos perjudiciales para la salud. En este punto, la IA puede utilizarse tanto como medio para proporcionar la información (*chatbots*), como asistir a los ciudadanos en el proceso de toma de decisiones a través de la realización de predicciones sobre las posibles consecuencias de sus actos. De este modo, se podría incidir en los beneficios y los costes de elegir ciertos comportamientos frente a otros. Por ejemplo, en el caso del pago de impuestos, haciendo visibles todas las posibles responsabilidades legales y problemas morales derivados del impago, así como el porcentaje de personas que son descubiertas realizando este tipo de infracciones. También se podrían hacer predicciones sobre las implicaciones positivas de pagar impuestos (porcentaje de personas sanadas a través de la seguridad social, número de niños escolarizados, etc.).

Tercero, otra forma de mejorar la precisión es a través de mecanismos que ayuden a los ciudadanos a planificar ciertos eventos que pueden ayudarles en su día a día. En este sentido las instituciones públicas podrían ser útiles para identificar situaciones en las que los ciudadanos necesitan un extra de ayuda para realizar una tarea. Por ejemplo, recordatorios para comer sano, pagar impuestos, etc. Aquí la IA puede ser especialmente útil para ayudar a los ciudadanos a planificarse de la manera más eficaz y eficiente a través de la identificación y conocimiento de sus hábitos, preferencias, experiencias previas, etc.

Finalmente, el último paso consiste en la evaluación del *nudging*, su eficacia y la introducción de posibles cambios de mejora (BIT, 2014). Al igual que en otros procesos de evaluación como el que tiene lugar en el ciclo de políticas públicas (Höchtel *et al.*, 2016; Valle-Cruz *et al.*, 2020), podemos esperar que la incorporación de la IA en este último paso del *nudging* tenga consecuencias similares. La principal ventaja asociada al uso de tecnologías de IA en los procesos de evaluación es la posibilidad de análisis y procesamiento de la información en tiempo real. De este modo, es posible detectar irregularidades o problemas que no deberían tener lugar en el momento que suceden. Esta nueva característica permitiría transformar radicalmente la práctica del *nudging* ya que el proceso de evaluación no tendría que quedar reservado para el último momento, sino que podría incorporarse en todos los pasos del proceso de implementación. Aunque este sería el mayor cambio asociado a la incorporación de IA, también se podrían observar otras oportunidades como una mayor precisión en la medición de los resultados del *nudging* y en la identificación de errores, entre otras cosas.

5. DISCUSIÓN

En la sección anterior hemos visto como la aplicación de tecnologías de IA, así como el aprovechamiento de sus distintas capacidades, puede no solo ayudar a mejorar la implementación y el uso de *nudges* en el sector público, sino realmente llegar a transformar este tipo de intervenciones dando lugar a lo que podríamos denominar políticas conductuales inteligentes y, más concretamente, *intelligent nudging*. Los *nudges* inteligentes tendrían niveles de eficacia y eficiencia más elevados que los de sus hermanos «no inteligentes» gracias al uso de tecnologías de reconocimiento facial, detección de objetos, procesamiento del lenguaje natural, predicción, análisis en tiempo real, etc., que permitirían mejorar todos los pasos del proceso.

A pesar del posible atractivo que pueden suponer las políticas conductuales inteligentes existen numerosos riesgos implicados en la puesta en marcha y funcionamiento de este tipo de políticas. Sin embargo, antes de abordar estos riesgos es importante señalar que la viabilidad del *intelligent nudging* hoy en día no solo es difícil por los problemas ético-políticos que implica, sino por el hecho de que el uso de IA en el sector público todavía no se encuentra suficientemente extendido y carece de una legislación específica (Misuraca y Van Noordt, 2020; Van Roy *et al.*, 2021). Si bien,

los esfuerzos realizados tanto por gobiernos como por administraciones públicas en toda Europa para adoptar tecnologías de IA y transitar hacia nuevos modelos de gobernanza (Comisión Europea, 2018) invita a pensar que el *intelligent nudging* podría convertirse en una realidad en el medio plazo. Por tanto, consideramos que vale la pena discutir los posibles riesgos ético-políticos, descritos anteriormente (Tabla 2), que se podrían derivar del uso de las políticas conductuales inteligentes.

Al igual que sucedía en el caso de los *nudges* tradicionales, en el caso del *intelligent nudging* los principales problemas ético-políticos están asociados con la disminución de las libertades individuales, la capacidad de decisión y elección de los ciudadanos, la manipulación y el determinismo (Bovens, 2009; Sunstein, 2020a). Además de estos problemas existen otros específicos derivados exclusivamente del uso de la IA y que actualmente están siendo discutidos en profundidad como la opacidad algorítmica, los problemas de privacidad, los sesgos, la falta de rendición de cuentas, etc. (Mittelstadt *et al.*, 2016; Taddeo y Floridi, 2018; Coeckelbergh, 2020). En relación con este segundo grupo de problemas, aquellos relacionados con la falta de privacidad son los más relevantes de cara a una puesta en marcha del *intelligent nudging*.

Los problemas de privacidad en las políticas conductuales inteligentes surgen de manera similar a los que se producen en otro tipo de cuestiones en las que se usa IA y se deben, principalmente, a la gran cantidad de datos que sería necesaria recopilar para poder llevar a cabo con éxito este tipo de políticas, así como la dificultad para anonimizar los datos de forma que sea imposible extraer o inferir ningún tipo de información sobre la persona de la que procede. Aunque es cierto que la recopilación de ciertos tipos de datos biométricos, así como otra de información sensible relativa a la orientación sexual de las personas y demás, no es una práctica legal ni ética, al menos dentro de la Unión Europea (Comisión Europea, 2021), la realidad es que desde hace décadas empresas y administraciones públicas de todo el mundo han estado recopilando grandes cantidades de datos que contienen información sensible acerca de la ciudadanía gracias a la falta de conocimiento y concienciación de la misma (Véliz, 2021).

Además, hoy en día sabemos que mucha información comprometida puede obtenerse no solo a través de la recogida directa de datos, sino también a través de distintas prácticas que permiten, por ejemplo, deducir la orientación sexual de una persona en base a sus preferencias musicales o cinematográficas (Véliz, 2021:32). En este contexto, la puesta en funcionamiento del *intelligent nudging* podría suponer un aliciente más para que organizaciones y empresas continuasen y extendieran sus prácticas de recogida e intercambio de datos personales. Problemas que deberían sumarse a los ya existentes y que hacen aún más necesario abordar y resolver los problemas de privacidad relacionados con la IA y las tecnologías basadas en datos.

La cuestión de la sustracción de datos personales y el intercambio de estos entre organizaciones estratégicas no solo es problemática de cara a la violación del derecho a la privacidad de la que goza la ciudadanía europea, sino que también guarda relación con otros problemas fundamentales como el de la libertad, la autonomía humana y el determinismo tecnológico. Algunos de estos problemas ya han sido señalados en el uso de *nudges* tradicionales. Por un lado, algunas críticas al *nudging* apuntan al excesivo

individualismo del que parten este tipo de políticas, olvidando mientras cuestiones relativas a la «práctica social» (Bicchieri, 2017; Güemes, 2021). A estas acusaciones se le suman otras como la supuesta manipulación que pueden generar las políticas conductuales o el riesgo de determinismo. Por otro lado, algunos liberales acusan a los defensores del *nudging* de paternalistas y señalan el exceso de carga moral de las políticas conductuales que buscan favorecer los «mejores» «adecuados» o «correctos» comportamientos. Siendo estas, según los críticos, cuestiones altamente subjetivas, dependientes de las preferencias personales y nunca asunto del Estado (Sunstein, 2020b).

A pesar de las críticas de los primero, los defensores del *nudging* no esconden el enfoque liberal e individualista del que parten y argumentan que es justo esto, el liberalismo que enfatiza la individualidad, lo que evita que los *nudges* produzcan dinámicas de control social y prácticas deterministas al mismo tiempo que logran producir cambios en los comportamientos grupales (Sunstein, 2014, 2020a). En este sentido, el presente artículo no entra a discutir si la Teoría del *nudging* es, realmente o hasta que grado, compatible con el liberalismo. Aun asumiendo que así fuese, en este caso nos centramos en los cambios producidos por la incorporación de la IA en los *nudges* que sí que podrían suponer un riesgo significativo para las democracias liberales más allá de la cuestión de la privacidad anteriormente mencionada. En este sentido, las políticas conductuales inteligentes podrían poner en riesgo no solo ciertas formas de libertad, sino especialmente la autonomía humana tal y como la entendemos en las liberal-democracias Occidentales.

En Occidente la idea de libertad vigente está muy vinculada a las reformulaciones del concepto de libertad que hicieron autores neoliberales como Milton Friedman (2022) o Friedrich Hayek (2022), entre otros, de los escritos liberales clásicos de John Locke (1988), Stuart Mill (2013), etc. En este sentido, la libertad está estrechamente relacionada con la no existencia de impedimentos para actuar siguiendo los designios de nuestra propia voluntad (Skinner, 1998; Bialy, 2022). Aunque la puesta en marcha de sistemas de IA, de por sí, podría no suponer un riesgo excesivo para este sentido de la libertad, las políticas conductuales inteligentes, por su alcance y naturaleza sí pueden resultar más invasivas a este respecto. Sin embargo, para poder comprender este riesgo es necesario desarrollar primero los peligros del *intelligent nudging* para la autonomía humana y las posibilidades de determinismo.

Al igual que en el caso de la libertad, los problemas relacionados con la autonomía humana y el determinismo ya estaban presente tanto en la literatura sobre *nudges* como en la de IA; sin embargo, estas cuestiones adquieren mayor magnitud al juntar ambos aspectos. La capacidad de análisis y predicción (no siempre acertada debido a sesgos en los datos, problemas en el diseño de los algoritmos, opacidad, etc.) de la IA, aplicada al ámbito de las políticas conductuales puede producir un efecto amplificador de los riesgos ya presentes. Por un lado, los *nudges* son sospechosos de reducir la autonomía humana y aumentar el determinismo al aprovechar nuestros sesgos cognitivos en beneficio de bien común (elegir comer ciertos alimentos saludables en vez de otros de menor calidad nutricional y así ahorrar al sistema de salud los costes que supondría

un potencial tratamiento para curar una enfermedad causada por una mala alimentación). Por otro lado, el uso de la IA puede ser difícil de compatibilizar con la autonomía humana (Taddeo y Floridi, 2018). Por ejemplo, cuando utilizamos una plataforma para consumir películas o series y elegimos el contenido que nos recomienda el algoritmo de la plataforma, estamos reduciendo nuestra autonomía humana al descartar (casi sin pensarlo) una gran cantidad de contenido que podría haber sido de nuestro agrado y generado modificaciones en nuestras preferencias. En estos casos, los usuarios deciden seguir un camino determinado en base a preferencias anteriores que sustituye la tarea activa de búsqueda y selección de contenido que se realizaba antes de la existencia de estas plataformas.

Aunque estos ejemplos puedan parecer anecdóticos, la combinación de ambos recursos (el *nudging* y la IA) sí puede generar escenarios potencialmente peligrosos para la autonomía humana. Si al hecho de que las políticas conductuales están enfocadas a alterar el comportamiento de la ciudadanía se le suman sistemas de IA, las capacidades de modificación de la conducta podría pasar de moderada a alta. Algo importante, no solo por las implicaciones en materia de derechos y libertades, sino porque los propios defensores del *nudging* con frecuencia aluden a su capacidad moderada para cambiar las dinámicas grupales cuando quieren defender su compatibilidad con la democracia liberal (Sunstein, 2014, 2020a). De este modo, lo que habría comenzado siendo un pequeño «empujón» en la dirección «correcta» (*nudge*), se convertiría en una incorporación forzosa a una autopista de una sola dirección con pocas salidas (*nudge* inteligente).

Por ejemplo, aunque algunos *nudges* y sistemas de IA, por separado, pueden utilizarse para reducir el abandono escolar o ayudar a los estudiantes a enfocar su futuro profesional en base a sus habilidades y capacidades más destacadas, estos no reducen necesariamente la autonomía humana. Sin embargo, las políticas conductuales inteligentes podrían, en base a perfiles personalizados realizados a través de herramientas de IA y aprovechando nuestros sesgos cognitivos, dirigirnos, a través de lo que parecería una simple sugerencia, hacia un camino que nosotros no hemos elegido de forma completamente autónoma. En este caso enfocado a la educación esto equivaldría, por ejemplo, a realizar recomendaciones personalizadas a ciertos colectivos para «invitarlos» a ocupar ciertos puestos de trabajo en el futuro.

En este sentido, la posible perversión de las políticas conductuales inteligentes y su principal riesgo sería el de generar escenarios en los que la ciudadanía tuviera importantes dificultades para salirse del camino establecido y quedasen fácilmente atrapados en senderos de dependencia. En el caso de las políticas conductuales inteligentes estos problemas son especialmente complejos porque, aunque parten de la idea de la voluntariedad y la posibilidad de no ser seguidos, existe el riesgo de que la ciudadanía no llegue a ser del todo consciente del determinismo al que estaría sometido. En este sentido, el *intelligent nudging* podría ser considerado un instrumento aparentemente amable, pero que realmente ejercería control social a través de una reducción de la libertad humana sin que nosotros apenas fuésemos conscientes de ello. Así, el riesgo de estas políticas reside en que los ataques a la libertad no son percibidos de forma

directa al poder la ciudadanía seguir realizando todo aquello que desea, sino de forma indirecta al disminuir su rango y capacidad de elección. Este tipo de políticas podrían jugar un papel esencial en el capitalismo de la vigilancia (Zuboff, 2020).

Además de estos riesgos, el *intelligent nudging* presenta otra serie de potenciales problemas como la presencia de sesgos y la discriminación que estos generan (De Zárate-Alcarazo, 2023). En este caso, por ejemplo, cualquier tipo de medida orientada a la selección de problemas, el análisis de comportamientos y la personalización de los *nudges*, podría presentar sesgos de género, raza, etc., como sucede en otros casos (Feine *et al.*, 2020). En las políticas conductuales inteligentes los sesgos podrían producirse tanto por un desconocimiento de los sesgos y la heurística específica de ciertos colectivos como por los datos y algoritmos de la IA. En la IA los sesgos se producen principalmente por la falta de representatividad en los datos. En este sentido, la presencia de sesgos es problemática porque podría sobredimensionar algunos problemas al mismo tiempo que invisibilizaría otros. Además, los sesgos en la identificación y análisis del contexto ya sea por error, o de manera intencionada, pueden alterar de forma importante la decisión de qué comportamientos deben cambiarse, así como el tipo de estímulos e información que se utiliza para dicha tarea. Todos estos elementos hacen que los problemas de manipulación ya presentes en el uso de *nudges* tradicionales se agudice al introducir la IA.

6. CONCLUSIONES

El creciente interés por la aplicación de políticas conductuales para solucionar los problemas de la ciudadanía, así como las intenciones de convertir la IA en un grupo de tecnologías decisivo en el sector público hacen que sea importante explorar cuales serían las implicaciones de la interacción entre ambos mundos. La IA ya por si sola promete aportar grandes beneficios si se adopta de manera ética y responsable. Lo mismo sucede en el caso de las políticas conductuales. De este modo, la combinación de la IA y los *nudges* podría convertirse en una buena oportunidad para mejorar la eficacia y la eficiencia este tipo de intervenciones y mejorar la vida de los ciudadanos.

Sin embargo, el uso de tecnologías de reconocimiento facial, de reconocimiento del entorno, predicción, etc., junto a los problemas asociados a la recopilación de información personal conllevan importantes retos ético-políticos que cuestionan seriamente la idoneidad del desarrollo de las políticas conductuales inteligentes. Las amenazas que el *intelligent nudging* puede suponer para la privacidad, la libertad y la autonomía humana, genera serias dudas sobre su compatibilidad con los principios liberales democráticos de Occidente. De este modo, las políticas conductuales inteligentes generarían aún más dudas que las ya producidas por su homólogos «analógicos» y, de no ser diseñadas, implementadas y usadas, de forma ética contribuirían directamente a ahondar en el capitalismo de la vigilancia.

Por tanto, el presente artículo supone un intento por esbozar las principales oportunidades que la IA podría suponer para el *nudging* y los riesgos ético-políticos

vinculados a la aparición de las políticas conductuales inteligentes. Si bien, de cara a profundizar en el estudio del *intelligent nudging* sería interesante centrarse en la búsqueda de aplicaciones reales de *nudges* inteligentes en el sector público para poder estudiar estas políticas de forma más detallada, cómo han sido diseñadas, sus beneficios y sus consecuencias. También es importante señalar la importancia de estudiar otros métodos alternativos al EAST en la aplicación de *nudges*, así como los problemas relacionados con sus limitaciones respecto al comportamiento colectivo (Güemes, 2021).

Referencias

- Almazán, Adrian. 2021. *Técnica y tecnología: Cómo conversar con un tecnólogo*. Madrid: Taugenitl.
- Behavioural Insights Team. 2012. «Applying Behavioural Insights to Reduce Fraud, Debt and Error» Disponible en web: https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/60539/BIT_FraudErrorDebt_accessible.pdf
[Consulta: 26 de mayo de 2023]
- Behavioural Insights Team. 2014. «EAST: Four Simple Ways to Apply Behavioural Insights». Disponible en web: https://www.bi.team/wp-content/uploads/2015/07/BIT-Publication-EAST_FA_WEB.pdf
[Consulta: 26 de mayo de 2023]
- Beshears, John, Choi, James J., Laibson, David, y Madrian, Brigitte C. 2009. «The Importance of Default Options for Retirement Saving Outcomes: Evidence from the United States» en Brown, Jeffrey, Liebman, Jeffrey y Wise, David A., *Social Security Policy in a Changing Environment*. Chicago: University of Chicago Press. <http://www.nber.org/chapters/c4539>
- Bettinger, Eric P., Long, Bridget T., Oreopoulos, Philip y Sanbonmatsu, Lisa. 2012. «The role of application assistance and information in college decisions: Results from the H&R block FAFSA experiment», *The Quarterly Journal of Economics*, 127(3): 1205-1242. <http://dx.doi.org/10.1093/qje/qjs017>
- Biały, Filip. 2022. «Freedom, silent power and the role of an historian in the digital age - Interview with Quentin Skinner», *History of European Ideas*, 48(7): 871-878. <http://dx.doi.org/10.1080/01916599.2021.2020970>
- Bicchieri, Cristina. 2017. *Norms in the Wild. How to diagnose, measure and change social norms*. New York: Oxford University Press. <https://doi.org/10.4000/oecnomia.2980>
- Bovens, Luc. 2009. «The Ethics of Nudge», enGrüne-Yanoff, Till y Ove Hansson, Sven (Eds.) *Preference Change: Approaches from Philosophy, Economics and Psychology*. Springer. http://dx.doi.org/10.1007/978-90-481-2593-7_10
- Burrell, Jenna. 2016. «How the machine ‘thinks’: Understanding opacity in machine learning algorithms», *Big data & society*, 3(1). <http://dx.doi.org/10.1177/2053951715622512>

- Coeckelbergh, Mark. 2020. *AI Ethics*. MIT Press
- Comisión Europea. 2018. Artificial Intelligence for Europe. Disponible en web: <https://eur-lex.europa.eu/legalcontent/EN/TXT/PDF/?uri=CELEX:52018DC0237&from=EN> [Consulta: 29 de mayo de 2023]
- Comisión Europea. 2019. Ethics guidelines for trustworthy AI. Disponible en web: <https://ec.europa.eu/digital-singlemarket/en/news/ethics-guidelines-trustworthy-ai> [Consulta: 29 de mayo de 2023]
- Comisión Europea. 2021. Proposal for a Regulation laying down harmonised rules on artificial intelligence (Artificial Intelligence Act). Disponible en web: https://eur-lex.europa.eu/resource.html?uri=cellar:e0649735-a372-11eb-9585-01aa75ed71a1.0008.02/DOC_1&format=PDF [Consulta: 29 de mayo de 2023]
- De Sousa, Wesley Gomes, de Melo, Elis Regina Pereira, Bermejo, Paulo Henrique De Souza, Farias, Rafael Araújo Sousa, Gomes, Adalmir Oliveira. 2019. «How and where is artificial intelligence in the public sector going? A literature review and research agenda». *Government Information Quarterly*, 36(4). <https://doi.org/10.1016/j.giq.2019.07.004>
- De Zárate-Alcarazo, Lucía. O. 2022. «Explicabilidad (de la inteligencia artificial)», *Eunomía. Revista en Cultura de la Legalidad*. <http://dx.doi.org/10.20318/eunomia.2022.6819>
- De Zárate-Alcarazo, Lucía. O. 2023. «Sesgos de género en la Inteligencia Artificial», *Revista de Occidente*, 502, (1).
- Di Prospero, Adam, Norouzi, Nojan, Fokaefs, Marios y Litoiu, Marin. 2017. «Chatbots as assistants: an architectural framework» en *Proceedings of the 27th Annual International Conference on Computer Science and Software Engineering*. <https://dl.acm.org/doi/10.5555/3172795.3172805>
- Ebeling, Felix, y Sebastian, Lotz. 2015. «Domestic Uptake of Green Energy Promoted by Opt-out Tariffs», *Nature Climate Change*, 5: 868-871. <http://dx.doi.org/10.1038/nclimate2681>
- Einfeld, Colette. 2019. «Nudge and evidence-based policy: fertile ground», *Evidence & Policy: A Journal of Research, Debate and Practice*, 15(4): 509-524. <http://dx.doi.org/10.1332/174426418X15314036559759>
- Esmark, Anders. 2019. «Communicative governance at work: how choice architects nudge citizens towards health, wealth and happiness in the information age», *Public Management Review*, 21(1): 138-158. <http://dx.doi.org/10.1080/14719037.2018.1473476>
- Ewert, Benjamin. 2020. «Moving beyond the obsession with nudging individual behaviour: Towards a broader understanding of Behavioural Public Policy», *Public Policy and Administration*, 35(3): 337-360. <http://dx.doi.org/10.1177/0952076719889090>
- Ewert, Benjamin., y Loer, Kathrin. 2021. «Advancing behavioural public policies: in pursuit of a more comprehensive concept», *Policy & Politics*, 49(1): 25-47. <https://doi.org/10.1332/030557320X15907721287475>
- Feine, Jasper, Gnewuch, Ulrich, Morana, Stefan y Maedche, Alexander. 2020. «Gender Bias in Chatbot Design» en *Chatbot Research and Design: Third International Workshop*. http://dx.doi.org/10.1007/978-3-030-39540-7_6

- Friedman, Milton. 2022. *Libertad de elegir: una declaración personal*. Madrid: Deusto.
- Gasser, Urs y Almeida, Virgilio A. F. 2017. «A Layered Model for AI Governance.» *IEEE Internet Computing*, 21 (6): 58-62. <http://dx.doi.org/10.1109/MIC.2017.4180835>
- Gigerenzer, Gerd. 2008. *Rationality for mortals: How people cope with uncertainty*. Oxford: Oxford University Press.
- Gigerenzer, Gerd. 2014. *Risk Savvy: How to make good decisions*. New York: Viking.
- Gopalan, Maithreti y Pirog, Maureen A. 2017. «Applying behavioral insights in policy analysis: Recent trends in the United States», *Policy Studies Journal*, 45(1): 82-114. <http://dx.doi.org/10.1111/psj.12202>
- Greenemeier, Larry. 2014. *Smart Machines Join Humans in Tracking Africa Ebola Outbreak* *Scientific American*. Disponible en web: <https://www.scientificamerican.com/article/smart-machines-join-humans-in-tracking-africa-ebola-outbreak/> [Consulta: 29 de mayo de 2023]
- Guan, Bin, Bao, Guoxian, Liu, Qingchuan y Raymond, Raduziner G. 2021. «Two-Way Risk Communication, Public Value Consensus, and Citizens' Policy Compliance Willingness About COVID-19: Multilevel Analysis Based on Nudge View», *Administration & Society*, 53 (7): 1106-1149.
- Güemes, Cecilia. 2017. «¿Nudge en América Latina? Incidir en comportamiento individual, obtener resultados colectivos», *Reforma y Democracia*, 68: 43-74.
- Güemes, Cecilia. 2021. Políticas conductuales. *Eunomía. Revista En Cultura de La Legalidad*, 20(20): 310-321. <http://dx.doi.org/10.20318/eunomia.2021.6078>
- Guihot, Michael, Matthew, Anne F. y Suzor, Nicolas P. 2017. «Nudging robots: Innovative solutions to regulate artificial intelligence», *Vanderbilt Journal of Entertainment & Technology Law*, 20(2).
- Hayek, Friedrich. 2022. *Los fundamentos de la Libertad*. Madrid: Alianza
- Hallsworth, Michael, List, John A., Metcalfe, Robert D. y Vlaev, Ivo. 2017. «The behavioralist as tax collector: Using natural field experiments to enhance tax compliance», *Journal of public economics*, 148: 14-31. <https://doi.org/10.1016/j.jpubeco.2017.02.003>
- Harguindéguy, Jean Baptiste. 2013. *Análisis de Políticas Públicas*. Madrid: Tecnos.
- Höchtel, Johann, Parycek, Peter y Schöllhammer, Ralph. 2016. «Big data in the policy cycle: Policy decision making in the digital era», *Journal of Organizational Computing and Electronic Commerce*, 26(1-2): 147-169. <http://dx.doi.org/10.1080/10919392.2015.1125187>
- Just, Natascha y Latzer, Michael. 2017. «Governance by algorithms: reality construction by algorithmic selection on the Internet», *Media, Culture and Society*, 39(2): 238-258. <http://dx.doi.org/10.1177/0163443716643157>
- Kahneman, Daniel y Tversky, Amos. 1984. «Choices, Values, and Frames», *American Psychologist*, 39(4): 341-350. <http://dx.doi.org/10.1037/0003-066X.39.4.341>
- Kahneman, Daniel. 2011. *Thinking, fast and slow*. Macmillan.

- Kasdan, David O. 2019. «Nudging the Neoliberal Agenda: Administrative Opportunities in the Deregulated State», *Public Administration Review*, 79(3): 439-442. <http://dx.doi.org/10.1111/puar.13045>
- Kuehnhanss, Colin R. 2019. «The challenges of behavioural insights for effective policy design». *Policy and Society*, 38(1): 14-40. <http://dx.doi.org/10.1080/14494035.2018.1511188>
- Quest, Lisa, Charrie, Anthony, du Croo de Jongh, Lucas y Roy, Subas. 2018. «The risks and benefits of using AI to detect crime», *Harvard Business Review*: 2-5.
- Lazer, David. 2015. «The rise of the social algorithm», *Science*, 348(6239): 1090-1091. <http://dx.doi.org/10.1126/science.aab1422>
- Locke, John. 1988. *Two Treatises on Government*. Cambridge: Cambridge University Press.
- Lune, Howard y Berg, Bruce L. 2017. *Qualitative research methods for the social sciences*. Pearson.
- Madrian, Brigitte C. y Shea, Dennis F. 2001. «The power of suggestion: Inertia in 401(k) participation and savings behavior», *The Quarterly Journal of Economics*, 116(4): 1149-1187. <http://dx.doi.org/10.1162/003355301753265543>
- Margetts, Helen y Dorobantu, Cosmina. 2019. «Rethink government with AI» *Nature*, 568(7751): 163-165. <http://dx.doi.org/10.1038/d41586-019-01099-5>
- Meijer, Albert y Grimmelhuijsen, Stephan. 2020. «Responsible and Accountable Algorithmization: How to Generate Citizen Trust in Governmental Usage of Algorithms» en Schuilenburg, Marc y Schuilenburg, Rik Peeters (Eds.), *The Algorithmic Society*. Routledge.
- Mehta, Ravi y Zhu, Rui J. 2009. «Blue or red? Exploring the effect of color on cognitive task performances», *Science*, 323(5918): 1226-1229. <http://dx.doi.org/10.1126/science.1169144>
- Mill, J. S. (2013). *Sobre la Libertad*. Alianza.
- Misuraca, Gianluca. y Van Noordt, Colin. 2020. «AI Watch - Artificial Intelligence in public services» Disponible en web: <https://publications.jrc.ec.europa.eu/repository/handle/JRC120399> [Consulta: 29 de mayo de 2023]
- Mittelstadt, Brent D., Allo, Patrick, Taddeo, Mariarosaria, Wachter, Sandra, y Floridi, Luciano. 2016. «The ethics of algorithms: Mapping the debate», *Big Data & Society*, 3(2). <http://dx.doi.org/10.1177/2053951716679679>
- Mols, Frank, Haslam, S. Alexander, Jetten, Jolanda y Steffens, Niklas K. 2015. «Why a nudge is not enough: A social identity critique of governance by stealth», *European Journal of Political Research*, 54(1): 81-98. <http://dx.doi.org/10.1111/1475-6765.12073>
- Moseley, Alice y Stoker, Gerry. 2015. «Putting public policy defaults to the test: The case of organ donor registration», *International Public Management Journal*, 18(2): 246-264. <http://dx.doi.org/10.1080/10967494.2015.1012574>
- Oliver, Adam. 2019. «Towards a new political economy of behavioral public policy». *Public Administration Review*, 79(6): 917-924. <http://dx.doi.org/10.1111/puar.13093>

- Pohl, Rüdiger. F. 2022. *Cognitive illusions: Intriguing phenomena in judgement, thinking and memory*. Routledge. <http://dx.doi.org/10.4324/9781003154730>
- Prainsack, Barbara. 2020. «The value of healthcare data: to nudge, or not?» *Policy Studies*, 41(5): 547-562. <http://dx.doi.org/10.1080/01442872.2020.1723517>
- Reimer, Daniel y Houmanfar, Ramona A. 2017. «Internalities and their applicability for organizational practices» *Journal of Organizational Behavior Management*, 37(1): 5-31. C <https://doi.org/10.1080/01608061.2016.1257969>
- Russell, Stuart y Norvig, Peter. 2016. *Artificial Intelligence: a modern approach*. Pearson Education.
- Schmidt, Andreas T. 2017. «The power to nudge» *American Political Science Review*, 111(2): 404-417. <http://dx.doi.org/10.1017/S0003055417000028>
- Simon, Herbert A. 1987. «Making Management Decisions: the Role of Intuition and Emotion», *Academy of Management Perspectives*, 1(1): 57-64. <http://dx.doi.org/10.5465/ame.1987.4275905>
- Simon, Herbert A. 1991. «Bounded Rationality and Organizational Learning», *Organization Science*, 2(1): 125-134. <http://dx.doi.org/10.1287/orsc.2.1.125>
- Simon, Herbert A. 1993. «Decision Making: Rational, Nonrational, and Irrational», *Educational Administration Quarterly*, 29(3): 392-411. <http://dx.doi.org/10.1177/0013161X93029003009>
- Simon, Herbert A. 1997. *Administrative Behavior: A study of Decision-Making Processes in Administrative Organizations*. The Free Press.
- Simon, Herbert A. 2000. «Bounded Rationality in Social Sciences», *Mind & Society*, 1: 25-39.
- Skinner, Quentin. 1998. *Liberty before liberalism*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/CBO9781139171274>
- Sousa Lourenco, Joana, Ciriolo, Emanuele, Rafael Rodrigues Vieira De Almeida, Sara y Troussard Xavier. 2016. Behavioural Insights Applied to Policy - European Report. Disponible en web: <https://publications.jrc.ec.europa.eu/repository/handle/JRC100146> [Consulta: 29 de mayo de 2023]
- Strabheim, Holger. 2020. «The Rise and Spread of Behavioral Public Policy: An Opportunity for Critical Research and Self-Reflection», *International Review of Public Policies*, 2(1): 115-128. <https://doi.org/10.4000/irpp.897>
- Strassheim, Holger. 2021. «Behavioural mechanisms and public policy design: Preventing failures in behavioural public policy» *Public Policy and Administration*, 36(2): 187-204. <http://dx.doi.org/10.1177/0952076719827062>
- Sun, Tara Q. y Medaglia, Rony. 2019. «Mapping the challenges of Artificial Intelligence in the public sector: Evidence from public healthcare», *Government Information Quarterly*, 36(2): 368-383. <http://dx.doi.org/10.1016/j.giq.2018.09.008>
- Sunstein, Cass R. 2014. *Why nudge?: The politics of libertarian paternalism*. New Haven: Yale University Press. <http://dx.doi.org/10.12987/9780300206920>
- Sunstein, Cass R. 2020a. *Behavioral Science and Public Policy*. Cambridge: Cambridge University Press. <http://dx.doi.org/10.1017/9781108973144>

- Sunstein, Cass R. 2020b. *Too Much Information*. Cambridge: MIT Press. <http://dx.doi.org/10.7551/mitpress/12608.001.0001>
- Sunstein, Cass R., Reisch, Lucia A. y Kaiser, Micha. 2019. «Trusting nudges? Lessons from an international survey» *Journal of European Public Policy*, 26(10): 1417-1443. <http://dx.doi.org/10.1080/13501763.2018.1531912>
- Sunstein, Cass R., y Thaler, Richard H. 2003. «Libertarian paternalism is not an oxymoron», *The University of Chicago Law Review*: 1159-1202. <http://dx.doi.org/10.2307/1600573>
- Taddeo, Mariarosaria y Floridi, Luciano. (2018). «How AI can be a force for good», *Science*, 361: 751-752. <http://dx.doi.org/10.1126/science.aat5991>
- Thomann, Eva. 2018. «Donate your organs, donate life! Explicitness in policy instruments» *Policy Sciences*, 51(4): 433-456. <http://dx.doi.org/10.1007/s11077-018-9324-6>
- Thompson, Sam, Michaelson, Juliet, Abdallah, Saamah, Johnson, Victoria, Morris, Damien, Riley, Kathy y Simms, Andrew. 2011. «Moments of change as opportunities for influencing behaviour: A report to the Department for Environment, Food and Rural Affairs». Disponible en web: [https://orca.cardiff.ac.uk/id/eprint/43453/1/MomentsofChangeEV0506FinalReportNov2011\(2\).pdf](https://orca.cardiff.ac.uk/id/eprint/43453/1/MomentsofChangeEV0506FinalReportNov2011(2).pdf) [Consulta: 29 de mayo de 2023]
- Tummers, Lars. 2019. «Public policy and behavior change» *Public Administration Review*, 79(6): 925-930. <http://dx.doi.org/10.1111/puar.13109>
- Turner, Karen, Alabi, Oluwafisayo y Race, Julia. 2020. «Nudging policymakers: a case study of the role and influence of academic policy analysis», *Journal of European Public Policy*, 27(8): 1270-1286. <http://dx.doi.org/10.1080/13501763.2020.1742774>
- Tversky, Amos y Kahneman, Daniel. 2011. «Rational choice and the Framing of Decisions», *The Journal of Business*, 59(4): 251-278. <http://dx.doi.org/10.1017/CBO9780511598951.011>
- Thaler, Richard H. y Sunstein, Cass R. 2003. «Libertarian paternalism», *American economic review*, 93(2): 175-179. <http://dx.doi.org/10.1257/000282803321947001>
- Thaler, Richard H. y Sunstein, Cass R. 2008. *Nudge. Improving decisions about health, wealth, and happiness*. New Haven: Yale University Press.
- Van Roy, Vincent, Rossetti, Fiammetta, Perset, Karine y Galindo-Romero, Laura. 2021. «AI Watch - National strategies on Artificial Intelligence: A European perspective». Disponible en web: <https://publications.jrc.ec.europa.eu/repository/handle/JRC122684> [Consulta: 29 de mayo de 2023]
- Valle-Cruz, David, Criado, Juan Ignacio, Sandoval-Almazán, Rodrigo y Ruvalcaba-Gomez, Edgar A. (2020). «Assessing the public policy-cycle framework in the age of artificial intelligence: From agenda-setting to policy evaluation», *Government Information Quarterly*, 37: 1-12. <http://dx.doi.org/10.1016/j.giq.2020.101509>
- Véliz, Carissa. 2021. *La Privacidad es Poder: Datos, vigilancia y libertad en la era digital*. Barcelona: Debate. <https://doi.org/10.15304/ges.3.8673>
- Visintin, Emilio P., Bonvin, Jean-Michel, Varone, Frederic, Butera, Fabrizio, Lovey, Max y Rosenstein, Emilie. 2021. «Can street-level bureaucrats be nudged to

- increase effectiveness in welfare policy?», *Policy & Politics*, 49(1): 121-139. <http://dx.doi.org/10.1332/030557320X15955051687823>
- Vlaev, Ivo, King, Dominic, Dolan, Paul y Darzi, Ara. 2016. «The theory and practice of «nudging»: changing health behaviors», *Public Administration Review*, 76(4): 550-561. <http://dx.doi.org/10.1111/puar.12564>
- Zauberman, Gal, Kim, B. Kyu, Malkoc, Selin A. y Bettman, James R. 2009. «Discounting time and time discounting: Subjective time perception and intertemporal preferences», *Journal of Marketing Research* 46(4): 543-556. <http://dx.doi.org/10.1509/jmkr.46.4.543>
- Zuboff, Shoshana. 2020. *La era del capitalismo de la vigilancia*. Barcelona: Paidós. https://doi.org/10.46583/scio_2021.20

Presentado para evaluación: 21 de marzo de 2022.

Aceptado para publicación: 27 de abril de 2022.

LUCÍA ORTIZ DE ZÁRATE ALCARAZO

lucia.ortizdezarate@uam.es

ORCID: 0000-0001-7775-4378

Investigadora predoctoral en Ética y Gobernanza de la Inteligencia Artificial en la Universidad Autónoma de Madrid (UAM) e investigadora asociada en el Círculo de Bellas Artes de Madrid. Es graduada en Física y Filosofía (Premio Extraordinario de Fin de grado y Premio Alumni-UCM 2018/2019) por la Universidad Complutense de Madrid (UCM). En el año 2021 fue incluida como miembro de la segunda edición de la lista «35 under 35» de futuros líderes en Inteligencia Artificial y Gobernanza Algorítmica por el Foro CIDOB-Santander. Además, es Visiting Fellow 2023/2024 en el Programa de Ciencia, Tecnología y Sociedad de la Harvard Kennedy School (Universidad de Harvard).