

Consideraciones éticas en torno al uso de tecnologías basadas en datos masivos en la UNED

(Ethical considerations on the use of massive data-based technologies in UNED)

José L. Aznarte

Vicerrectorado de Digitalización e Innovación, UNED (España)

DOI: <http://dx.doi.org/10.5944/ried.23.2.26590>

Cómo referenciar este artículo:

Aznarte, J. L. (2020). Consideraciones éticas en torno al uso de tecnologías basadas en datos masivos en la UNED. *RIED. Revista Iberoamericana de Educación a Distancia*, 23(2), pp. 237-252. doi: <http://dx.doi.org/10.5944/ried.23.2.26590>

Resumen

En este documento se dibuja una posición ética respecto al uso de tecnologías basadas en datos masivos, con atención particular al contexto de los procesos de enseñanza/aprendizaje en la UNED. Se aboga aquí por una toma de conciencia de los riesgos concretos que van aparejados con las ventajas esperables del uso de dichas tecnologías. Además, se esbozan algunas preguntas importantes y algunas respuestas posibles, incluyendo nueve cautelas esenciales, a fin de proporcionar unas bases sólidas a los proyectos que impliquen el uso de datos masivos en la UNED. Primero se expone lo que entendemos por una ética del cuidado, en la que la prioridad sea cuidar del alumnado a través del uso de datos y al mismo tiempo tener cuidado de que cualquier intervención basada en datos sea pulcra en sus presunciones y no introduzca sesgos. A continuación, se expone una selección de referentes y aproximaciones previas que han sido tenidas en cuenta y se enumeran las preguntas clave que deben guiar a cualquier institución que se plantee el trabajo con datos. Finalmente, se enuncian nueve cautelas esenciales que la institución se compromete a observar y que han sido propuestas a la comunidad universitaria para su deliberación.

Palabras clave: ética; recogida de datos; análisis estadístico; inteligencia artificial.

Abstract

This document draws up an ethical position about the use of massive data-based technologies is drawn, with a particular focus on the teaching and learning processes in UNED. We intend to raise awareness on the concrete risks associated with the potential advantages of using those technologies, and we set out some important questions and some possible answers to them, including nine essential cautions which can provide a sound basis for any projects requiring the use of massive data in UNED.

In the first place, we explain what we understand by an ethics of care, in which the priority must be to take care of the students through the use of their data and, at the same time, to care about any data-based intervention being clean in its presumptions and bias-free. Then we consider a selection of previous works that have inspired ours, and we enumerate key questions over which any institution must ponder before starting to work with personal data. Finally, we state nine cautions that our institution adheres to and that every member of the University has been invited to deliberate upon.

Keywords: ethics; massive data; statistical analysis; artificial intelligence.

El uso de datos para la toma de decisiones ha sido sometido a un desarrollo exponencial desde que una parte importante de nuestras comunicaciones, nuestro entretenimiento, nuestras relaciones comerciales y nuestros aprendizajes suceden en la Web. Las compañías que proporcionan servicios en Internet han liderado el desarrollo de técnicas de extracción de información a partir de los grandes volúmenes de datos que genera nuestra presencia en la red, a fin de, por ejemplo, identificar consumidores potenciales de sus productos, para refinar esos mismos productos con vistas a incrementar su rentabilidad o para dirigir sus iniciativas de marketing hacia las preferencias individuales de clientes potenciales. Más recientemente, la investigación y el desarrollo de sistemas educativos *online* ha emprendido la exploración de estas mismas técnicas con un objetivo distinto: el de mejorar los procesos de enseñanza/aprendizaje.

Así, cuando las tecnologías basadas en datos masivos se aplican a los procesos de enseñanza/aprendizaje, conllevan el desarrollo de métodos y la aplicación de técnicas de la estadística y la inteligencia artificial que, sumados a conocimientos de las ciencias de la educación, de la sociología y de la psicología, entre otros, permiten analizar los datos recolectados durante dichos procesos. La minería de datos educativos, por un lado, se centra en la búsqueda de patrones en los datos y en el desarrollo de nuevos algoritmos o modelos a fin de descubrir información que ayude a explicar mejor los procesos educativos. A su vez, las analíticas de aprendizaje buscan la aplicación de algoritmos o modelos existentes que permitan la creación de aplicaciones que influyeran directa y positivamente los procesos de enseñanza/

aprendizaje. Todo ello se enmarca en el progreso hacia escenarios de aprendizaje personalizados.

El desarrollo y la implantación de estas tecnologías en la educación superior se ha iniciado principalmente durante las dos últimas décadas, antecedido por un desarrollo de más de 40 años de la inteligencia artificial, y se espera que su evolución sea muy considerable en el futuro cercano. Muchas universidades en el mundo están actualmente explorando sus posibilidades y la UNED no es una excepción. Sin embargo, su rápida adopción no ha sido generalmente acompañada de una reflexión sosegada acerca de las consideraciones éticas en torno a su uso, por lo que son muchas las voces que alertan de los riesgos y las amenazas que ello supone (Roberts et al., 2017).

En la UNED se ha abordado la implantación de estas tecnologías a gran escala con cierto retraso respecto a nuestro entorno (un ejemplo pionero que ya está en marcha es el proyecto de Inteligencia Institucional, concebido para optimizar los procesos de toma de decisión, planificación estratégica y rendición de cuentas). Sin embargo, esto, lejos de suponer un problema, supone precisamente la oportunidad de construir sobre cimientos mejor fundados las posibles aplicaciones que harán uso de unas tecnologías que, probablemente, cambiarán la forma en que esta institución presta su servicio a la comunidad.

Ejemplos de aplicaciones en el ámbito educativo

Del amplio conjunto de aplicaciones desarrolladas o en uso en la actualidad, las que siguen son algunas de las más relevantes dentro del ámbito de los procesos de enseñanza/aprendizaje. Un resumen más detallado, junto a referencias a trabajos científicos que muestran evidencias de lo dicho, puede encontrarse, por ejemplo, en el informe *Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics*, del Departamento de Educación de los Estados Unidos de América (Bienkowski et al., 2012):

- **Modelado del conocimiento del alumnado:** se trata de responder, con base en los datos, a la pregunta de qué sabe una persona concreta en términos de habilidades específicas, conceptos, conocimientos procedurales o capacidades superiores de razonamiento. Esta técnica ha sido adoptada para construir contenidos adaptativos, sistemas de recomendación o sistemas de tutorización inteligentes en los que el modelo de alumnado sirve para elegir qué contenidos son mostrados en cada caso, por ejemplo. La idea es usar los datos para adaptar el comportamiento del sistema a las necesidades específicas de cada persona, de forma que se emitan los mensajes correctos en el momento adecuado y de la forma más útil.
- **Modelado del comportamiento del alumnado:** busca responder a preguntas acerca del significado de los patrones de comportamiento en los procesos de

aprendizaje y acerca de la motivación y el compromiso del alumnado. A partir de datos como el tiempo que pasa conectado a la plataforma, si ha completado ciertas asignaturas, posibles cambios en el contexto educativo o el modelo de conocimiento disponible, se trata de hallar correlaciones entre el comportamiento y el éxito en el aprendizaje. El objetivo es dar respuesta a preguntas como: ¿una mayor interacción conlleva un mejor aprendizaje? ¿Qué patrones de interacción se relacionan con el abandono?

- **Modelado de la experiencia del alumnado:** consiste en averiguar si el alumnado está satisfecho con su experiencia y puede abordarse desde encuestas, desde los cuestionarios de satisfacción, pero también desde el análisis de decisiones, comportamiento, desempeño y probabilidad de abandono en temas o asignaturas subsiguientes. Particularmente interesante es el uso de esta información para el rediseño de cursos a fin de minimizar el abandono y maximizar la satisfacción del alumnado.
- **Segmentación del alumnado:** consiste en el proceso de clasificación del alumnado en perfiles definidos por un conjunto de datos personales que describen las características de cada individuo. Una vez que han sido clasificados, es posible predecir su comportamiento o sus preferencias a fin de mejorar su desempeño mediante entornos de aprendizaje adaptados a cada perfil.
- **Modelado de contenidos:** es posible extraer, de forma automática, un modelo del contenido de un tema o de una asignatura, identificando las unidades de estudio o los conceptos clave y sus relaciones al tiempo que se genera una taxonomía del dominio y de las asociaciones entre las competencias, respuestas del alumnado y acciones programadas sobre recursos educativos.
- **Diseño de procesos de aprendizaje:** se trata de analizar los datos de los distintos componentes de una metodología de enseñanza, en varios momentos y con muestras seleccionadas de estudiantes, para responder a cuestiones como: ¿qué elementos son más efectivos para promover el aprendizaje? ¿un cambio concreto en los contenidos puede mejorar el desempeño y la retención? ¿Qué tipo de metodologías son mejores en términos de la experiencia del alumnado?

Todas estas aplicaciones están en desarrollo y pueden ser tenidas en cuenta para su aplicación en el ámbito de procesos de enseñanza/aprendizaje. Sin embargo, es importante señalar que, obviamente, los datos pueden ser usados también en otros ámbitos y que en el futuro pueden surgir nuevas propuestas relacionadas con el uso de datos masivos. Por ese motivo, si bien está centrado en los procesos de enseñanza/aprendizaje, el marco ético que se propone en este documento aspiran a englobar cualquier otro uso de datos masivos que pueda ser concebido en el futuro, esté o no relacionado con dichos procesos.

HACIA UNA ÉTICA DEL CUIDADO EN EL USO DE DATOS

La misma ley para el león y el ratón es tiranía.
–William Blake–

La ética de los datos ha sido definida como una nueva rama de la ética que estudia y evalúa los problemas morales relacionados con los datos (su generación, registro, curado, procesado, diseminación, compartición y uso), los algoritmos (inteligencia artificial, aprendizaje automático, robots...) y las prácticas correspondientes (innovación responsable, programación, hackeo y códigos profesionales) a fin de formular y promover soluciones moralmente buenas (conductas o valores correctos) (Floridi et al., 2016).

El término 'dato' proviene del latín *datum*, que es aquello que es dado. Sin embargo, puesto que en el contexto tecnológico actual la mayor parte de los datos son tomados antes que dados, quizá sería más lógico hablar de *captum*: aquello que es apresado o capturado. Esta digresión sirve para contextualizar que los datos no existen independientemente de las ideas, los cuerpos, los instrumentos, las prácticas, los contextos y los conocimientos usados para generarlos, procesarlos y analizarlos (Kitchin, 2014). Así, el uso de las tecnologías basadas en datos debe ser visto como un nudo de motivaciones sociales, políticas, económicas y culturales que están plagadas de complicaciones, contradicciones y conflictos (Selwyn, 2013), y por tanto es nuestra responsabilidad el hacernos las preguntas adecuadas acerca de nuestra comprensión del potencial y los desafíos que comporta la recolección, el análisis y el uso de datos de estudiantes y profesorado.

En este sentido, en un contexto en el que se dan relaciones de poder, relaciones asimétricas, entre estudiantes, profesorado e institución académica, no cabe descartar que la percepción de la justicia en el uso de los datos esté determinada, en el nivel institucional, por consideraciones no directamente asumibles por los demás estamentos. Así, es esencial complementar la ética de la justicia en el uso de los datos con una ética del cuidado que implique una comprensión *relacional* de la recolección, análisis y uso de los datos y que al mismo tiempo ofrezca resistencias y alternativas a los procesos de vigilancia institucional.

Así, una ética del cuidado en el uso de datos (Prinsloo et al., 2017) propone activar el gesto y la mirada de la cautela: *tener cuidado* con nuestras presunciones acerca de las evidencias que aportan los datos, así como las que hacemos acerca de la naturaleza y el uso de estos, acerca de su gobernanza, de la privacidad y del acceso a los mismos. Una tal ética del cuidado, en una universidad pública, enraíza también con la noción de *cuidar del alumnado* y recopilar, analizar y utilizar sus datos para mejorar sus tasas de éxito y de retención. Finalmente, una ética del cuidado debe considerar que, aunque los intereses de la institución deben confluír con los del alumnado (como define el Real Decreto 2310/72 de creación de la UNED), de hecho,

puede ocurrir que no se dé esta coincidencia, y también que es necesario priorizar, en caso de conflicto, estos últimos.

REFERENTES Y APROXIMACIONES PREVIAS

Conscientes de las delicadezas éticas derivadas de su implantación, algunas instituciones educativas pioneras en el uso de tecnologías basadas en datos masivos se han preocupado por dotarse de marcos conceptuales, éticos y normativos para su uso. Y otras iniciativas han fracasado justamente por no hacerlo.

Es conocido el ejemplo de InBloom, una corporación financiada con más de 100 millones de dólares por las fundaciones Gates y Carnegie en 2011, cuyo objetivo era digitalizar, almacenar y utilizar los datos del alumnado de todo el sistema educativo norteamericano para, en teoría, avanzar hacia una educación personalizada mediante el uso de tecnologías basadas en datos. Sin embargo, preocupaciones acerca de la privacidad de los datos y, sobre todo, el uso comercial que de ellos podrían hacer terceras empresas a través de acuerdos económicos con InBloom alentaron una rebelión de asociaciones de padres y madres (*Student Privacy Matters*, 2019) que resultaron en la cancelación del proyecto en su totalidad en 2014 (Singer, 2014).

Entre las instituciones universitarias que más han avanzado en considerar las implicaciones éticas y sociales del uso de datos está la Open University del Reino Unido que, ya en 2014, publicó una normativa sobre el «Uso ético de los datos del alumnado para la analítica de aprendizaje» (Open University, 2019). Además, ha publicado diferentes versiones divulgativas para los miembros de la comunidad universitaria, empezando por un documento en el que se exponen los ocho principios clave en que se basa. Estos principios se agrupan en cuatro ámbitos (alinear el uso de datos con los principios de la Universidad, establecer claramente el propósito y los límites, implicar al alumnado en el uso de sus datos y asegurar que estos son usados de forma sensata) y han sido criticados por, entre otras cosas, no ser suficientemente exhaustivos en su redacción (Farrow, 2015).

Otra experiencia relevante tiene origen también en el Reino Unido: se trata del Código para la Práctica de las Analíticas de Datos publicado por Jisc en 2015 (Bailey y Scatter, 2015). Jisc (anteriormente conocida como *Joint Information Systems Committee*) es una organización sin ánimo de lucro que representa los intereses de las instituciones de educación superior del Reino Unido en lo tocante a servicios y soluciones digitales. Su código, concebido para ser aplicado de forma general en cualquier institución educativa, aspira a establecer las responsabilidades de dichas instituciones para asegurar que las tecnologías basadas en datos son utilizadas de forma responsable, apropiada y efectiva.

Otra aproximación al mismo asunto fue publicada en 2015 en el marco del proyecto europeo LEA's BOX (Steiner et al., 2016). Tras analizar en profundidad otras propuestas previas, estas son sintetizadas en un marco de privacidad y protección de datos con ocho requisitos fundamentales (privacidad, propósito y propiedad de los

datos, consentimiento, transparencia y confianza, acceso y control, responsabilidad y evaluación, calidad de los datos y seguridad y manejo de los datos). Este enfoque preconiza que las tecnologías basadas en datos en el marco de los procesos de enseñanza/aprendizaje deben alinearse con estos ocho principios, que serían, así, requisitos fundamentales para un código de conducta aceptable.

En 2016, en el marco del proyecto europeo *Learning Analytics Community Exchange* (LACE), que conjugó el trabajo de nueve socios de siete países, se publicó una lista de comprobación (Drachsler y Greller, 2016) para establecer las características necesarias para otorgar confianza a las aplicaciones de tecnologías basadas en datos a la educación superior. Esta lista, llamada DELICATE por sus siglas en inglés, se plantea como un compromiso que se compone de ocho puntos: determinación (de la necesidad del uso de datos), explicación (de objetivos y límites), legitimación (de que es apropiado el uso de datos), implicación (de todos los agentes), consentimiento (mediante un contrato), anonimización (que proteja la privacidad de los individuos), aspectos técnicos (para garantizar la privacidad) y colaboradores externos (que han de asumir enteramente el compromiso).

Otra experiencia muy relevante es la del proyecto SHEILA (Tsai et al., 2018), cofinanciado por el programa Erasmus+ de la UE, cuyo objetivo es ayudar a las universidades europeas a convertirse en usuarios y custodios más maduros de los datos masivos de estudiantes que aprenden *online*. La principal aportación de este proyecto es un marco de desarrollo de políticas institucionales que promueve la evaluación formativa y el aprendizaje personalizado a través de la implicación directa de los miembros de la comunidad en el proceso de desarrollo. Este marco está además soportado por una útil herramienta interactiva¹ que está basada en un repositorio de experiencias colectivas de 89 instituciones de 26 países distintos.

Ya en 2019, el Consejo Internacional para la Educación Abierta y a Distancia (ICDE por sus siglas en inglés), ha publicado un conjunto de directrices sobre las prácticas éticas en la aplicación en el ámbito educativo de tecnologías basadas en datos (Slade y Tait, 2019). El objetivo del documento es identificar asuntos centrales de relevancia global, entre los que se encuentran la transparencia, la propiedad y el control de los datos, su accesibilidad, su validez y fiabilidad, la responsabilidad institucional, las comunicaciones, los valores culturales, la inclusión, el consentimiento o la agencia y responsabilidad del alumnado.

Finalmente, es importante inscribir estas consideraciones éticas sobre el uso de tecnologías basadas en datos educativos en el marco de investigación e innovación responsable (*responsible research & innovation*, RRI) promovido por la Comisión Europea en el programa Horizonte 2020 (Comisión Europea, 2014).

CUESTIONES RELEVANTES EN EL USO DE TECNOLOGÍAS BASADAS EN DATOS MASIVOS

Cualquier aplicación de las tecnologías basadas en datos masivos, y particularmente aquellas que se propongan en el ámbito educativo, debe responder a una serie de preguntas relevantes, como las propuestas por Roberts et al., (2017), que adaptamos aquí. Siendo que muchas no tienen respuesta evidente, fácil o incluso posible, estas preguntas han de ser tenidas en cuenta en cada caso y acompañar cada decisión de diseño en la posible implantación de dichas tecnologías, también en ámbitos distintos del educativo.

- ¿Quién puede beneficiarse?

Más allá de categorías genéricas como “la comunidad universitaria”, es esencial considerar qué agentes concretos pueden verse beneficiados en cada caso, y en qué condiciones. En algunas ocasiones el beneficio de una de las partes puede ser incompatible con el de otras. Por supuesto, en tales casos, atendiendo a la ética de los cuidados considerada más arriba, se debe priorizar a las partes a priori más débiles. Además, es importante tener en cuenta las consecuencias que sobre aquellas personas que no se podrán beneficiar de estas tecnologías (alumnado en reclusión forzosa, con diversidad funcional o quienes no deseen que sus datos sean usados, por ejemplo) puede tener su generalización.

- ¿A quién puede poner en riesgo?

Además de considerar los beneficiarios, es importante identificar aquellos colectivos que pueden ser puestos en riesgo para poder calibrar los pros y los contras de cualquier intervención. Como es obvio, el riesgo será mayor para los colectivos más expuestos, que en el caso de la UNED son el alumnado y, en menor medida, los equipos docentes y tutoriales.

Para el primer colectivo, hay encuestas que indican que las tecnologías basadas en datos masivos suelen ser bien aceptadas, pero también que existen casos en que los efectos pueden ser perniciosos: por ejemplo, los mensajes reiterativos pueden inducir desmotivación en ciertas personas o las predicciones de abandono funcionar como profecías autocumplidas. Además, allí donde se usen datos demográficos y de formación previa, existe un potencial para la propagación de estereotipos, sesgos y discriminaciones basadas en características grupales.

En cuanto a los equipos docentes y tutoriales, si las preguntas que guían el uso de las tecnologías basadas en datos masivos son fijadas por la administración educativa, sin aportes de dichos equipos, es fácil que se incremente el control gerencial sobre la enseñanza, con el foco centrado en la rendición de cuentas. En este

sentido, el compromiso y la participación del profesorado es crucial para asegurar la implementación exitosa de cualquier aplicación.

- ¿Quién toma las decisiones?

Dada la preocupación por el uso inadecuado de los datos, es necesario identificar en cada caso dónde reside la responsabilidad de fijar las preguntas que han de ser respondidas mediante el análisis de los datos, pero también qué datos se recolectan, analizan y visualizan, y quién tendrá acceso a qué información. Experiencias previas recomiendan que la comunidad afectada (el alumnado y los equipos docentes y tutoriales, por ejemplo) sean parte del proceso decisorio, es decir, que no existan procesos meramente informativos, sino que sean vinculantes: que estén preparados para virar el rumbo si así lo decide la comunidad. Además, deberían establecerse procedimientos formalizados que vayan más allá de recoger información mediante encuestas o participación en foros, por ejemplo, utilizando metodologías de previsión de impacto o ingeniería de requisitos basada en la comunidad (Maalej et al., 2016).

- ¿Qué datos se ponen en juego?

Con la digitalización de sucesivamente más ámbitos de la vida cotidiana, incluyendo los procesos de enseñanza/aprendizaje, el volumen y la variedad de datos que pueden ser puestos en juego es cada vez mayor. Al mismo tiempo, las posibilidades de cometer errores y propagar sesgos se incrementan cuando los datos son parciales, no existentes o cuando se cruzan datos entre bases de datos diferentes. Las inexactitudes, sesgos e imprecisiones derivadas del proceso de tratamiento de los datos pueden tener consecuencias graves, por lo que se recomienda el acceso al mayor número de datos posibles. También es extremadamente delicado el uso potencial de fuentes de datos externas a la institución como las redes sociales, en las que se dan procesos de desvelamiento de identidad potencialmente incontrolados, que además descansan sobre políticas de compañías no necesariamente alineadas con los principios de la institución.

- ¿Quién necesita ser informado y qué necesita saber?

En el caso concreto de los procesos de enseñanza/aprendizaje, analógicos o digitales, es claro que dependen en todo caso de cierto nivel de cesión de información entre las partes. Esto es así también en otros ámbitos de digitalización. Sin embargo, el desarrollo de tecnologías basadas en datos masivos exige preguntarse hasta qué punto las instituciones tienen la obligación de informar a la comunidad del uso que se hace de sus datos y de que su comportamiento puede estar siendo hasta cierto punto monitorizado. Por ejemplo, ¿es necesario informar al alumnado de que se

almacenan datos anónimos para calcular promedios de tiempo empleado en acabar una titulación? ¿Y si por el contrario se almacenan los datos de personas concretas para el mismo fin?

- ¿Es necesario el consentimiento expreso para el uso de datos? ¿Cómo se otorga?

Más allá de informar de que hay datos que están siendo registrados, existe un debate abierto acerca de si el consentimiento expreso es necesario o no, y en caso de que sí, con qué frecuencia y en qué circunstancias. Algunos enfoques propuestos incluyen: aceptación de inclusión al inicio de la relación con posibilidad de decidir la autoexclusión más adelante; posibilidad de autoexclusión al inicio de la relación con posibilidad de inclusión más adelante; aceptación de inclusión al inicio de la relación y nueva aceptación de inclusión cada vez que haya cambios; aceptación de inclusión o posibilidad de exclusión cada vez que se desencadene una intervención basada en datos. La lista no es exhaustiva y cada enfoque tiene sus propias limitaciones.

- ¿Hasta qué punto son necesarias disposiciones de privacidad?

La privacidad, en el contexto actual, se ha convertido en un concepto polémico y resbaladizo. En este contexto, surgen preguntas acerca de qué datos se recopilan, quién puede ver información acerca de personas concretas (por ejemplo, ¿puede el profesorado ver información sobre el alumnado de cursos en los que no está involucrado?, pero también, ¿pueden otras instituciones?, ¿y futuros empleadores?) y cuánto tiempo es imprescindible mantener los datos accesibles, sabiendo que el uso de datos históricos de cursos pasados puede beneficiar al alumnado actual.

- ¿Es necesario desidentificar los datos? ¿Es esto suficiente para garantizar el anonimato?

Cuando el identificador personal es sustituido por un código arbitrario, decimos que los datos han sido desidentificados, y se espera que sea imposible reconocer a un individuo concreto a partir de los registros guardados. Sin embargo, cuando se juntan fuentes de datos diversas y detalladas ('big data'), esta precaución no siempre es suficiente, lo cual ha provocado, en algunos casos célebres, consecuencias inesperadas (por ejemplo, es posible identificar a una persona en una base de datos anonimizada mediante información de redes sociales). Estas consecuencias pueden ser evitadas con procesos de anonimización más robustos, pero, sin embargo, no siempre es posible (ni deseable) anonimizar datos durante el desarrollo de la relación entre la institución y los miembros de su comunidad, lo que supone, por ejemplo, que en ámbitos educativos los procesos de anonimización han de tener lugar una vez que el alumnado abandona la institución (o cuando fije la normativa).

- ¿Quién es responsable de la administración y la protección de los datos?

Es importante (y en algunos casos obligatorio) que la institución defina quién tiene la responsabilidad de la preservación, seguridad y compartición de los datos, incluyendo responsables concretos para la recolección, los procesos de anonimización, los análisis y la administración de los datos, al menos.

- ¿Quién posee los datos y cuáles son las implicaciones?

Determinar los derechos de propiedad intelectual de los datos y los resultados obtenidos a partir de ellos es un asunto de gran complejidad y repercusiones: ¿pertenecen a cada individuo, a la institución, a la compañía que provee la infraestructura? ¿Pueden usarse en investigación? ¿Cómo afecta esto a la posibilidad de reutilización de datos fuera de la UNED, si tal cosa llega a ser planteada?

- ¿Tiene la UNED alguna obligación de actuar?

Si se demuestra que las tecnologías basadas en datos masivos conllevan beneficios, como por ejemplo incrementar el éxito del alumnado, ¿tiene la institución la obligación de intervenir para sacar partido de esta oportunidad? Esta obligación, ¿qué relación guarda con el coste económico que dicha intervención pueda tener? Es decir, ¿es admisible que la UNED no actúe a sabiendas de que el alumnado puede beneficiarse, por ejemplo?

- ¿Cómo puede ser reconocida y respetada la voluntad de los miembros de la comunidad universitaria?

Si las tecnologías basadas en datos masivos sugieren alguna acción concreta, es necesario definir si las personas afectadas están o no obligadas a aceptarla, y cuáles son las consecuencias de una u otra elección. Existe un cierto riesgo de infantilización a través de la coerción y las invasiones de la privacidad.

- *Si algo sale mal, ¿qué posibilidades de reparación existen?*

Reconociendo que las tecnologías basadas en datos masivos pueden tener consecuencias imprevistas, es necesario preguntarse acerca de las capacidades de reparación de las que la institución quiere dotarse. Dar la posibilidad de que cualquiera pueda formular objeciones o quejas por un canal institucional claro es un paso esencial para abordar repercusiones inesperadas.

- ¿No está todo esto cubierto por la normativa vigente?

Pese a que la reforma europea de 2016 de las reglas de protección de datos (Unión Europea, 2016) ha supuesto un gran avance en la protección del control individual sobre los datos personales y en la especificación de reglas para instituciones y empresas, no faltan voces críticas que alertan de sus limitaciones en la protección de la privacidad (Biosca, 2018). Además, son notorios los casos de empresas que, cumpliendo escrupulosamente con la legalidad, han incurrido en prácticas más que cuestionables. Mediante este documento, la UNED pretende ir más allá y establecer un marco ético que, dentro del más escrupuloso respeto por la normativa vigente, proporcione mayor seguridad y tranquilidad a los miembros de su comunidad.

CAUTELAS NECESARIAS PARA UN USO RESPONSABLE DE TECNOLOGÍAS BASADAS EN DATOS MASIVOS

La UNED, atendiendo a lo anteriormente expuesto y consciente de su responsabilidad a la hora de prevenir los riesgos asociados con el uso de datos masivos en el contexto educativo y en otros, se compromete a observar, en todas las fases de diseño, implantación y uso de tecnologías basadas en datos, al menos las siguientes cautelas:

- **Cautela 0: Del cuidado**

Una voluntad expresa de cuidado guiará en todo momento las actuaciones de la UNED respecto al uso de los datos: tendrá cuidado de que no se haga un mal uso de ellos (*primum nil nocere*) y cuidará, a través de ellos, del bienestar de los miembros de la comunidad universitaria, particularmente del de los colectivos más vulnerables.

- **Cautela 1: De la responsabilidad**

La UNED asume su responsabilidad en el uso de datos y determinará una autoridad reconocible que será responsable del uso legal, ético y eficiente de las tecnologías basadas en datos. Además, será necesario fijar quién tiene responsabilidades específicas sobre la anonimización de los datos, así como su recolección, conservación y administración.

- **Cautela 2: De la transparencia**

La UNED desarrollará políticas institucionales claras con respecto al uso de tecnologías basadas en datos masivos. Definirá, registrará y comunicará a la comunidad universitaria las fuentes de datos, los propósitos de los análisis, las

métricas usadas, quién tiene acceso a los análisis, los límites de su uso y cómo se interpretan los datos. Cuando los datos estén incompletos o se usen como aproximación a otros no disponibles, se aclarará en qué suposiciones se basa dicha aproximación.

- **Cautela 3: Del consentimiento**

El consentimiento de cada miembro de la comunidad universitaria será necesario para el uso de sus datos personales. Si el consentimiento se da al inicio de la relación con la UNED, será informado con explicaciones detalladas acerca del uso previsto de los datos. Cuando los datos se usen para intervenir en las decisiones que afectan al trayecto de una persona en el seno de la institución o al acceso a sus recursos, será necesario obtener consentimiento expreso y específico para esos usos. La UNED establecerá protocolos para permitir la revocación del consentimiento.

- **Cautela 4: De la privacidad y el acceso**

El acceso a los datos y a los análisis derivados de ellos estará restringido a quienes tengan una causa legítima, que será determinada por la UNED en función del nivel de agregación de los datos y las competencias de quienes lo soliciten. Cuando los datos sean anónimos, la UNED cuidará de que no sea posible reidentificar a los individuos a partir de los metadatos ni por agregación de múltiples fuentes de datos. La UNED tendrá particular cuidado, en el caso de que los datos sean cedidos a terceras partes, de que estas se adhieran a las cautelas aquí expresadas y a los principios de la institución, evitando particularmente usos comerciales.

- **Cautela 5: De la propiedad y el control**

La UNED asume que no es la propietaria de los datos personales recabados, sino solo la responsable temporal de su tratamiento. Conforme a las prácticas indicadas por la Agencia Española de Protección de Datos en cumplimiento del Reglamento Europeo de Protección de Datos, la UNED cuidará de que los miembros de su comunidad tengan la posibilidad de corregir, eliminar o añadir contexto a sus datos siempre que sea posible, así como la de acceder a los análisis derivados y la de reclamar ante posibles consecuencias adversas del uso de tecnologías basadas en datos.

- **Cautela 6: De la validez y la fiabilidad**

Para asegurar que las aplicaciones de tecnologías basadas en datos son válidas y fiables, la UNED garantizará que los datos son precisos y representativos de aquello

que dicen medir. Además, serán mantenidos al día tanto como sea posible. Cuando en el análisis se usen encuestas de opinión o se apliquen inferencias estadísticas, la UNED cuidará de que la muestra sea suficientemente grande y representativa y de que los resultados sean estadísticamente significativos. Todos los algoritmos y métricas utilizados serán comprendidos, validados, revisados y mejorados según corresponda por personal calificado.

- **Cautela 7: De los posibles impactos adversos**

La UNED reconoce que cualquier individuo es siempre más que la suma de los datos disponibles acerca de ella o él, y que las circunstancias personales no pueden ser descritas totalmente por los datos. Así, tomará medidas para evitar que tendencias, promedios, categorías o etiquetas produzcan sesgos en la percepción de la institución sobre las personas o en su relación con ellas, así como que se refuercen actitudes discriminatorias o se incrementen las desigualdades. Además, el impacto de las intervenciones basadas en datos sobre los distintos colectivos de la comunidad universitaria será tenido en cuenta, particularmente en las necesidades de formación y en la carga de trabajo. En cualquier caso, la UNED hará lo que esté en su mano por minimizar los posibles impactos adversos.

- **Cautela 8: De la participación**

Siempre que sea posible, la UNED tratará de involucrar a los distintos colectivos de la comunidad universitaria en la aplicación de tecnologías basadas en datos. En particular, dadas las relaciones asimétricas de poder en relación con el alumnado, otros colectivos de la UNED tratarán a sus miembros como iguales en lo relativo a los usos de sus datos, cuidando de que su punto de vista sea tenido en cuenta en la toma de decisiones.

NOTAS

- ¹. Este documento tiene por objeto proporcionar contexto y material para el diseño de un proceso participativo para deliberar en torno a las cautelas definidas en la Sección 5. Dicho proceso participativo tuvo lugar del 16 al 26 de enero de 2020 a través de la web <http://participa.uned.es>. Las cautelas que resultaron de dicho proceso compondrán un documento institucional que será asumido por la UNED.

Se agradece la contribución de Jesús González Boticario, Miguel Santamaría Lancho, Ángeles Manjarrés Riesco, Inés Gil Jaurena, José Ramón Álvarez Sánchez, Juan Manuel Lacruz López, Mariano Melendo Pardos, Belén Ruiz de Gopegui, Ángeles Sánchez-Elvira Paniagua, Patricia Sánchez Sánchez del Arco, Ricardo Mairal Usón, Paz San Segundo Manuel, Gomer Betancor Nuez y Javier Gil García.

REFERENCIAS

- AEPD (2019). *Guía práctica para las evaluaciones de impacto en la protección de los datos sujetas al RGPD*. Agencia Española de Protección de Datos. Recuperado de <https://www.aepd.es/media/guias/guia-evaluaciones-de-impacto-rgpd.pdf>
- Bailey, P. y Scletter, N. (2019). Code of practice for learning analytics, *Jisc*. Recuperado de <https://www.jisc.ac.uk/guides/code-of-practice-for-learning-analytics>
- Bienkowski, M., Feng, M., y Means, B. (2012). *Enhancing teaching and learning through educational data mining and learning analytics: An issue brief*, U.S. Department of Education, Office of Educational Technology.
- Biosca, P. (2018). Lo mejor y lo peor de la nueva normativa de privacidad, según los expertos, *ABC*. Recuperado de https://www.abc.es/tecnologia/redes/abci-rgpd-mejor-y-peor-nueva-normativa-privacidad-segun-expertos-201805242219_noticia.html
- Blake, W. (1790). *The Marriage of Heaven and Hell*. En *Erdman, David V. (Ed.), The Complete Poetry and Prose*. Versión de José Viñals.
- Comisión Europea (2014). Responsible research & innovation, *Horizon 2020 - European Commission*. Recuperado de <https://ec.europa.eu/programmes/horizon2020/en/h2020-section/responsible-research-innovation>
- Drachler H., y Greller W. (2016). Privacy and Analytics: It's a DELICATE Issue a Checklist for Trusted Learning Analytics, en *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, New York, NY, USA, (pp. 89-98).
- Farrow, R. (2015). Ethical principles of learning analytics – mini critique. Recuperado de <https://philosopher1978.wordpress.com/2015/12/03/ethical-principles-of-learning-analytics-mini-critique/>
- Floridi, L., y Taddeo, M. (2016). What is Data Ethics?, *Philosophical Transactions of the Royal Society A*, 374(2083).
- Kitchin, R. (2014). *The Data Revolution: Big Data, Open Data, Data Infrastructures & Their Consequences*. SAGE Publications Ltd.
- Maalej, W., Nayebi M., Johann, T., y Ruhe, G. (2016). Toward Data-Driven Requirements Engineering, *IEEE Software*, 33(1), 48-54.
- Open University (2019). *Ethical use of Student Data for Learning Analytics – Student Policies and Regulations* – Open University. Recuperado de <https://help.open.ac.uk/documents/policies/ethical-use-of-student-data>
- Prinsloo, P., y Slade, S. (2017). Big Data, Higher Education and Learning Analytics: Beyond Justice, Towards an Ethics of Care. En B. Kei Daniel, (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice*, (109-124). Cham: Springer International Publishing,
- Roberts, L. D., Chang, V., y Gibson, D. (2017). Ethical Considerations in Adopting a University- and System-Wide Approach to Data and Learning Analytics. En B. Kei Daniel (Ed.), *Big Data and Learning Analytics in Higher Education: Current Theory and Practice* (89-108). Cham: Springer International Publishing.
- Selwyn, N. (2013). *Distrusting educational technology: Critical questions for changing times*. Routledge.
- Singer, N. (2014). InBloom Student Data Repository to Close, *Bits Blog*. Recuperado de <https://bits.blogs.nytimes.com/2014/04/21/inbloom-student-data-repository-to-close/>
- Slade, S., y Tait, A. (2019). Global guidelines: Ethics in Learning Analytics, International

- Council for Open and Distance Education (ICDE). Recuperado de <https://www.icde.org/s/Global-guidelines-for-Ethics-in-Learning-Analytics-Web-ready-March-2019.pdf>
- Steiner, C. M., Kickmeier-Rust, M. D., y Albert, D. (2016). LEA in Private: A Privacy and Data Protection Framework for a Learning Analytics Toolbox, *Journal of Learning Analytics*, 3(1), 66-90.
- Student Privacy Matters (2019). Parent Coalition for Student Privacy, *Parent Coalition for Student Privacy*. Recuperado de <https://www.studentprivacymatters.org/>
- Tsai, Y. S. et al. (2018). SHEILA: Supporting Higher Education to Intergrade Learning Analytics. Recuperado de <https://sheilaproject.eu/2018/11/30/sheila-final-research-report/>
- Unión Europea (2016). *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos) (Texto pertinente a efectos del EEE)*, vol. 119. Recuperado de <http://data.europa.eu/eli/reg/2016/679/oj/spa>

PERFIL ACADÉMICO Y PROFESIONAL DEL AUTOR

José L. Aznarte. Profesor titular en el Departamento de Inteligencia Artificial de la UNED. Sus líneas de investigación giran en torno a la predicción de series temporales y sus aplicaciones para el bien común: por ejemplo, el modelado de la contaminación. Coordina el desarrollo de un sistema predictivo operacional para predecir la calidad del aire en la ciudad de Madrid. Es director de la cátedra UNED-EMT en Movilidad Sostenible y Calidad del Aire. Temporalmente ejerce como vicerrector adjunto de Digitalización e Innovación, donde trabaja en la aplicación de tecnologías basadas en datos masivos a los procesos de enseñanza/aprendizaje, con especial atención a las implicaciones sociales y éticas del uso de dichos datos. <https://sites.google.com/view/jlznarte>
E-mail: jlznarte@dia.uned.es

Dirección:
Departamento Inteligencia Artificial | UNED
C/ Juan del Rosal, 16. Ciudad Universitaria, 28040 Madrid, España
tel.: +34 913989688

Fecha de recepción del artículo: 28/01/2020
Fecha de aceptación del artículo: 13/02/2020
Fecha de aprobación para maquetación: 23/03/2020