

Ciencia Ciudadana

Guía Inteligencia Artificial



Fran Castillo



MINISTERIO DE EDUCACIÓN Y FORMACION PROFESIONAL

Dirección General de Evaluación y Cooperación Territorial Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado (INTEF) Recursos Educativos Digitales

La **Aventura de Aprender** es un espacio de encuentro e intercambio en torno a los aprendizajes para descubrir qué prácticas, atmósferas, espacios y agentes hacen funcionar las comunidades; sus porqués y sus comos o, en otras palabras, sus anhelos y protocolos. Este proyecto parte de unos presupuestos mínimos y fáciles de formular. El primero tiene que ver con la convicción de que el conocimiento es una empresa colaborativa, colectiva, social y abierta. El segundo abraza la idea de que hay mucho conocimiento que no surge intramuros de la academia o de cualquiera de las instituciones canónicas especializadas en su producción y difusión. Y por último, el tercero milita a favor de que el conocimiento es una actividad más de hacer que de pensar y menos argumentativa que experimental. Estas guías didácticas tienen por objetivo favorecer la puesta en marcha de proyectos colaborativos que conecten la actividad de las aulas con lo que ocurre fuera del recinto escolar. Sin aventura no hay aprendizaje, ya que las tareas de aprender y producir son cada vez más inseparables de las prácticas asociadas al compartir, colaborar y cooperar.

<http://laaventuradeaprender.intef.es/>

Proyecto concebido y coordinado por **Antonio Lafuente** para **INTEF**



Obra publicada con licencia de Creative Commons Reconocimiento-Compartir Igual 4.0
Licencia Internacional. <https://creativecommons.org/licenses/by-sa/4.0/>

Derechos de autor

Licencia CC BY-SA 4.0

Para cualquier asunto relacionado con esta publicación contactar con:

Instituto Nacional de Tecnologías Educativas y de Formación del Profesorado

C/Torrelaguna, 58. 28027 Madrid.

Tfno.: 91-377 83 00 | Fax: 91-368 07 09

Correo electrónico: lada@educacion.gob.es



Índice

Introducción.	04
¿Por qué la automatización no es suficiente en un escenario de Grandes Datos e Inteligencia Artificial?	04
Materiales para ejecutar el proyecto.	06
Paso a paso	07
Paso 4.0. Visualización y Pensamientos Impensables.	07
Paso 4.1. Un problema global.	10
Paso 4.2. Datos.	10
Paso 4.3. Algoritmos.	11
4.3.1. Aprendizaje supervisado.	11
4.3.2. Aprendizaje no supervisado.	12
4.3.3. Aprendizaje por refuerzo.	13
4.3.4. Redes neuronales convolucionales.	13
4.3.5. Redes neuronales recurrentes.	14
Paso 4.4. Una herramienta para diseñar inteligencias artificiales abiertas.	15
Paso 4.4.1. Nuevos patrones de diseño para la transparencia de datos.	15
Paso 4.4.2. Explicabilidad.	18
Paso 4.4.3. La inteligencia artificial comete errores.	21
Paso 4.4.4. Confianza y algoritmos.	23
Paso 4.4.5. Inteligencia artificial auditable.	25
Casos de uso.	26
Productos de datos e inteligencias aumentadas en un escenario energético y parkings inteligentes.	26
Principios de Diseño para APPS de rastreo de contactos afectados por COVID-19.	28
Consejos.	31
Recursos.	32

¿Quién hace esta guía?

Fran Castillo es consultor independiente en el ámbito de la inteligencia artificial y el producto digital, actualmente dirige *data-driven futures*. Durante los últimos años ha trabajado como *data innovation manager*, liderando el desarrollo de nuevos servicios & productos digitales en el ámbito del *big data* y la inteligencia artificial. Ha trabajado, entre otras, para compañías como *Synergic Partners* (Telefónica), *The Cocktail* (WPP) y *Accenture*.

Introducción

Cuando oímos el término de Inteligencia Artificial, ¿qué imagen, o pensamiento, viene a nuestra cabeza?

El deseo de esta guía es compartir un nuevo territorio que nos permita explorar, entender, y diseñar, otras formas de inteligencia artificial (IA).

En la actualidad, los datos y los algoritmos de IA ya están condicionando la forma en la que pensamos, vemos y nos relacionamos con el mundo. En este nuevo escenario, vamos a necesitar nuevos patrones de diseño, nuevas interfaces, que nos faciliten una nueva relación con los algoritmos y, en última instancia, con el mundo.

La guía pretende presentar diferentes conceptos, y casos de uso, que nos permitan comprender, primero, cómo funcionan los algoritmos de Inteligencia Artificial, pero sobre todo, facilitar un marco desde el que inspirar nuevas visiones en la

construcción de tecnologías de Inteligencia artificial más humanas.

El objetivo de la guía, por tanto, es presentar una herramienta que nos ayude, a todos, a pensar en nuevas direcciones en el diseño de tecnologías de IA que puedan extender, y amplificar, nuestra relación con los sistemas digitales en los que estamos, ya todos, inmersos.

Los algoritmos, que modelan nuestras vidas, deberían de ser objetos observables, abiertos, sobre los que poder, colectivamente, discutir las implicaciones de los efectos que imponen. El hecho de poder ser objetos observables nos facilitará construir nuevas formas de relación.

Durante la guía, os presentaremos diferentes ideas que nos ayuden a transformar estos algoritmos de inteligencia artificial en objetos observables.

¿Por qué la automatización no es suficiente en un escenario de Grandes Datos e Inteligencia Artificial?

En este nuevo escenario de grandes volúmenes de datos y algoritmos de aprendizaje automático (*machine learning*), el diseño de soluciones digitales está dominado, en la actualidad, por el paradigma de la automatización, escenario que hace referencia a la construcción de software que automatiza procesos, acciones, sin la necesidad de la intervención de un humano. Diseña, por tanto, mecanismos que posibilitan acelerar los procesos de toma de decisión y en el que lo relevante es la eficacia, y precisión, de los modelos (algoritmos).

Este enfoque es altamente extendido en el contexto actual, de la *economía de la atención*, donde ésta es, cada vez, más escasa. Estamos expuestos a gran cantidad de información y de ahí, la necesidad de construir soluciones digitales que faciliten o aceleren los procesos de toma de decisión.

La mayoría de nosotros sufrimos una fatiga cognitiva debido a la cantidad de decisiones que debemos tomar a diario. Además, nuestro cerebro manifiesta disonancias y sesgos cognitivos que

requieren de la necesidad de tomar atajos (heurísticas) para reducir el coste energético asociado a la toma de decisiones.

El enfoque actual de la industria de Grandes Datos (*Big Data*) y la Inteligencia Artificial es la construcción de productos de datos que puedan estar orientados a **simplificar el número de opciones y reducir el número de decisiones.**

En este contexto de la economía de la atención y la automatización como una estrategia de la industria para reducir las fricciones y maximizar la eficacia, ha surgido un concepto, en el ámbito del diseño, que es el *Diseño para Menos Opciones* (*Design for Less Choice*) o el *Diseño Anticipatorio* (*Anticipatory Design*) que propone Aaron Shapiro.

Aaron lo describe como el “*Design That’s One Step Ahead of You* -el diseño que está un paso por delante de ti-”. Ambos son enfoques del diseño de producto orientados a reducir el número de decisiones a tomar y por tanto, soluciones que van a requerir una menor carga cognitiva asociada.

En definitiva, diseños que reducen la forma en la que interaccionamos con un determinado sistema (por ejemplo, una ciudad, un sistema natural, el aire, la energía, etc.). Conceptos que se basan principalmente en las capacidades de los algoritmos y la analítica predictiva para determinar de forma anticipada las necesidades del usuario.

Imaginad, por ejemplo, ¿Cómo se traduciría esta idea del Diseño para un Menor Número de Opciones, o el Diseño Anticipatorio, en el contexto de Netflix?

Netflix, como sabéis, es un producto de datos que hace uso de algoritmos de aprendizaje automático para construir interfaces, recomendaciones y experiencias de usuario personalizadas en función de los patrones de comportamiento, historial y preferencias del usuario. Pero, ¿Cómo evolucionaría su interfaz y diseño en el contexto del Diseño Anticipatorio (Anticipatory Design)? En este escenario, y gracias a sus bases de datos enriquecidas con datos de terceros, podrán llegar a ser mucho más precisos en sus recomendaciones, reduciendo el número de películas recomendadas y explicitando... “Esta es la película o serie que deberías ver ahora atendiendo a tus gustos y cómo fue tu día”—*«This is the movie you should watch right now, according to your tastes and how your day was going...»*

Este tipo de soluciones digitales están basadas, como describíamos, en la relevancia de lo automático. Su propuesta de valor diferencial reside en la capacidad e inteligencia de sus algoritmos y no en el diseño de una experiencia de usuario, de un producto o interfaz, que nos permita tener una mayor comprensión de nuestros hábitos y patrones de comportamiento para poder tomar mejores decisiones.

La promesa del *Diseño Anticipatorio*, o el *Diseño para un Menor Número de Opciones*, es nuevamente la eliminación de la fricción y el aumento de la eficacia para mejorar los resultados y métricas de la compañía. Pero también, es un escenario, que va a simplificar la forma de relación con estas tecnologías de inteligencia artificial.

Estas soluciones operan en el paradigma de la estética de la eficacia y de la facilidad de uso. Los algoritmos, ya, automatizan e interpretan la realidad por nosotros, simplificando nuestra relación con las tecnologías, por lo tanto, delegando en las máquinas la toma de decisiones.

En este nuevo escenario de grandes volúmenes de datos, en el que ya estamos inmersos, nos gustaría poder explorar otras preguntas que puedan extender el paradigma de la automatización:

¿Qué otras interfaces, creéis, podemos diseñar que nos permitan entender mejor nuestros patrones de comportamiento, para tomar mejores decisiones y que no resida en el poder del algoritmo para decidir por nosotros?

¿Qué nuevas interfaces podemos diseñar que amplifiquen nuestra consciencia y conocimiento sobre el funcionamiento de los sistemas que modelan nuestras vidas?

¿Qué nuevos principios de diseño nos van a permitir una nueva relación con los datos que generamos?

Marina Garcés describe en su libro *“Nueva Ilustración Radical”*, si la utopía solucionista que propone E. Morozov ya no trata de ampliar las capacidades humanas, sino de lo que verdaderamente trata es de delegar la inteligencia misma en las máquinas. La inteligencia artificial, entendida así, propone Garcés, es una inteligencia delegada.

En esta guía, exploraremos un nuevo enfoque que nos permita entender la inteligencia artificial como un nuevo medio para amplificar nuestras capacidades cognitivas y extender nuestra forma de relación con el mundo.

Materiales para ejecutar el proyecto.

Para desarrollar el proyecto sugerido en esta guía, harán falta los siguientes materiales:

- Ordenador.
- Acceso a internet.
- Necesitarás papel y lápiz, o un dispositivo, como *tablet* o móvil, con un bloc de notas digital.
- En el proyecto a desarrollar, necesitaréis definir un problema global a resolver y exploraremos cómo la Inteligencia artificial puede ayudaros a solucionarlo.

Algunos ejemplos, de problemas a resolver, podrían ser:

¿Cómo podríamos diseñar una herramienta que nos permitiese mejorar nuestros patrones de consumo energético, que nos hiciese más conscientes del consumo eléctrico? ¿Cómo los datos y los algoritmos de inteligencia artificial pueden ayudarnos a conseguirlo?

¿Cómo podríamos diseñar una herramienta que nos ayudase a mejorar los niveles de contaminación de nuestras ciudades?

¿Cómo podríamos diseñar una herramienta de rastreo de contactos de enfermos COVID, atendiendo a nuevos, esquemas y, patrones de diseño que nos permitan un tratamiento de los datos más transparente y confiable?

¿Podríamos editar el algoritmo de Tinder para conocer a chicas más afines a nuestros gustos? ¿Cómo creéis que, actualmente, Tinder genera nuestros perfiles de afinidad? ¿Creéis que el algoritmo que han desarrollado puede modelar el amor?

¿Creéis que podríamos editar en Google Maps alguna funcionalidad que nos permitiese descubrir nuevos lugares donde salir? ¿Cómo deberían de ser estos lugares? ¿Qué nuevas métricas de eficacia podríamos utilizar, que vayan más allá de la variable tiempo? ¿Podríamos rediseñar Google Maps para optimizar la exploración, el descubrimiento de esos lugares invisibilizados donde encontrar a gente con nuestros gustos y no, solamente, atendiendo a la distancia mínima

entre dos puntos?

¿Podríamos conectar nuestras preferencias de música en Spotify con nuestras preferencias de películas y series en Netflix para encontrar nuevas recomendaciones sobre libros que leer o conciertos a los que poder asistir ...?

¿Por qué no podrían los navegadores oscurecerse cuando pasamos demasiado tiempo en internet o, incluso, cuando abrimos demasiadas ventanas? A veces se cierran, pero no por este motivo, porque alguien en Google, Firefox o Apple no previó nuestro exceso de información.

¿Por qué no crear otras formas de visualización que nos permitan entender cómo Google, por ejemplo, personaliza nuestros resultados en las búsquedas?

¿Cómo podríamos medir el impacto medioambiental de un algoritmo de aprendizaje profundo (deep learning, es, solamente, un tipo de algoritmo de inteligencia artificial)? ¿Cómo creéis que impacta en la naturaleza el despliegue de los principales algoritmos que dominan el mundo (Facebook, Google, Amazon, etc.). ¿Cómo sería una inteligencia artificial sostenible con el medio ambiente (Green AI)? ¿Qué forma creéis que tendría?

Y para finalizar, ¿Creéis que podría tener sentido que el algoritmo de Whatsapp o Instagram, pudiese identificar cuando estás junto a tus amigos y, en estas situaciones, impedirte acceder a internet?

El objetivo de estas preguntas no es más que inspirar diferentes tipologías de problemas, siéntete libre de seleccionar cualquier problema global que te pueda resultar de mayor interés.

- El proyecto propone, también, desarrollar un pequeño ejercicio complementario, en el que podamos entender mejor cómo los datos, que compartimos, modelan los servicios digitales que proveen las compañías tecnológicas. Necesitaréis, por tanto, seleccionar un servicio digital o una aplicación que utilicéis con mucha frecuencia en vuestro móvil.

Algunos ejemplos de apps podrían ser: Spotify, Instagram, TikTok, etc.

Paso a paso:

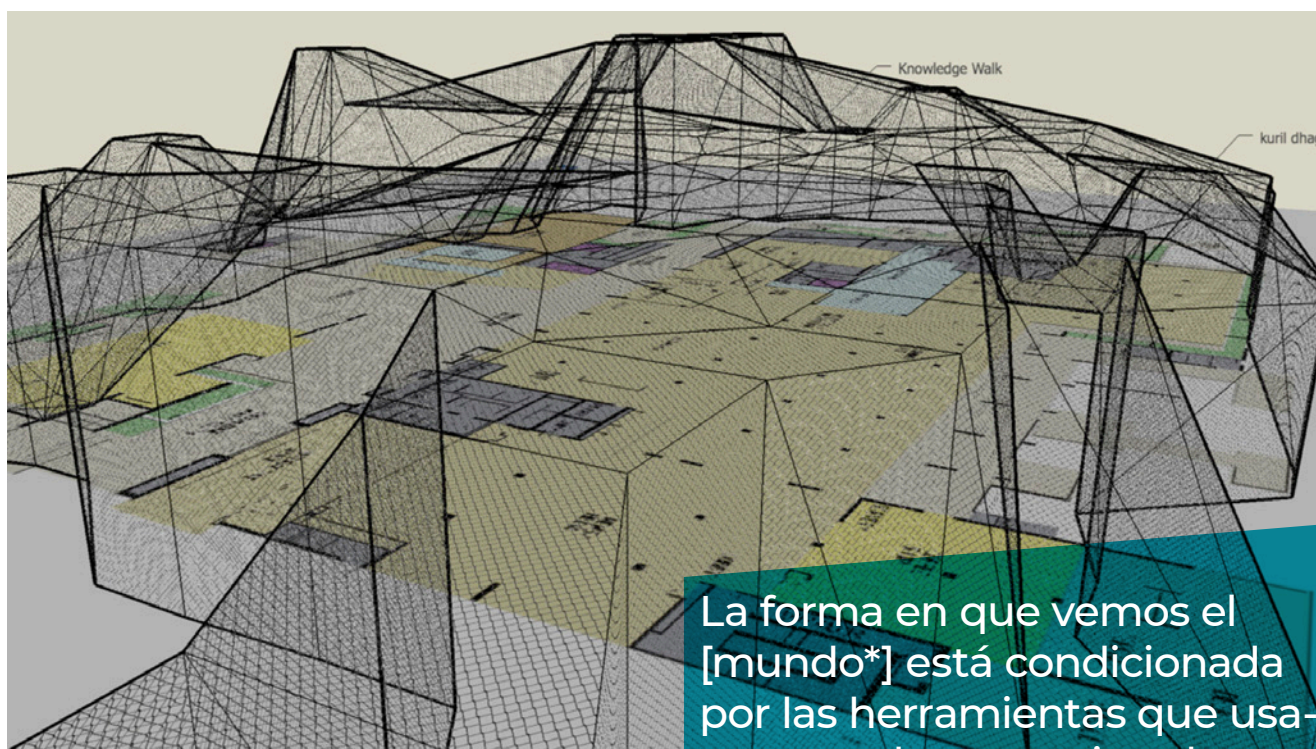
Antes de comenzar a pensar en términos de datos, algoritmos y nuevos principios de diseño en la construcción de Inteligencias Artificiales más “humanas”, me gustaría compartir algunas ideas que exploran la importancia de desarrollar

herramientas que nos permitan nuevas formas de ver en los datos, nuevas formas de ver, que posibilitaran otras formas de relación con los sistemas en los que estamos inmersos.

Paso 4.0. Visualización y Pensamientos Impensables *[Unthinkable Thoughts].*

¿Por qué, creéis, puede ser relevante diseñar nuevas interfaces, herramientas de visualización, que nos permitan nuevas formas de comprensión y exploración de la realidad y no solo como un medio de representación de la misma?

¿Cómo pensáis que la tecnología y los datos están cambiando la forma en la que vemos la realidad?



(Imagen Dan Hill – Wifi Structures). Licencia [CC BY-SA 4.0](#)

La imagen representa una visualización de las ondas Wifi en un entorno doméstico. Herramientas que posibilitan la representación de una cartografía de lo invisible, de espacios de datos no perceptibles, pero que claramente influyen en nuestros modos de relación.

La forma en que vemos el [mundo*] está condicionada por las herramientas que usamos, por lo tanto, si podemos diseñar nuevas herramientas que nos permitan cambiar la forma en la que vemos una determinada realidad, podremos cambiar el [mundo*]

*mundo es un término que podréis sustituir por cualquier otro.

El objetivo en esta guía, como os decía, es compartir algunas ideas que nos permitan entender la Visualización como un medio de exploración de la complejidad y no, solamente, como un sistema representacional, que es el enfoque más extendido en la actualidad. Creemos que estos enfoques son claramente limitantes para explorar y extraer nuevos *insights* de los datos.

Es en la fricción, y no en la eficacia, ni en la facilidad de uso, donde creemos existe un recurso productivo que va permitir poner en relieve complejidades que son muy difíciles de percibir.

"Just as there are odors that dogs can smell and we cannot, as well as sounds that dogs can hear and we cannot, so too there are wavelengths of light we cannot see and flavors we cannot taste.

Why then, given our brains wired the way they are, does the remark "Perhaps there are thoughts we cannot think," surprise you?

Evolution, so far, may possibly have blocked us from being able to think in some directions; there could be unthinkable thoughts."

Richard Hamming

Traducción cita Richard Hamming:

"Al igual que existen olores que los perros pueden oler y nosotros no, sonidos que los perros pueden oír y nosotros no, también existen longitudes de onda de luz que no podemos ver y sabores que no podemos distinguir.

¿Por qué entonces, dada la forma en la que están conectados nuestros cerebros, tiene sentido, que quizás haya pensamientos que no podemos pensar, te sorprende esto?

La evolución, hasta ahora, posiblemente nos haya impedido pensar en algunas direcciones; podría, entonces, haber pensamientos impensables."

Esta es una cita de Richard Hamming que ha sido muy inspiradora durante gran parte de mi carrera profesional. Hamming propuso, al igual que hay ciertos olores, sonidos y espectros de onda que no podemos percibir y otros animales si, él propuso, si tenía sentido plantear que puede haber pensamientos que no han podido ser pensados, y si la evolución nos ha impedido pensar en determinadas direcciones.



Imagen 1668 Anton Van Leeuwenhoek by Vermeer. Licencia [CC BY-SA 4.0](#)

Claro que existen estos Pensamientos impensables (*Unthinkable Thoughts*), pero históricamente hemos construido herramientas que nos han permitido amplificar nuestras capacidades

cognitivas, ampliar nuestros sentidos y de ahí, la primera idea que quería compartir, y es la Visualización como un nuevo medio para posibilitar estos pensamientos impensables.

Anton Van Leeuwenhoek, padre de la biología molecular, ideó lupas de aumentación que permitían visualizar una realidad biomolecular que hasta ese momento era imperceptible. Herramientas que nos permitieron nuevas relaciones con los sistemas biológicos. Por tanto, en este nuevo escenario de Grandes Datos, ¿Cuáles, creéis, son estas nuevas herramientas que nos van a permitir entender mejor la complejidad de los datos?, ¿Cuáles son estas nuevas herramientas que nos permiten representar la realidad de otras formas?

Este es un ejemplo, fantástico, para entender la visualización como una herramienta de exploración y de construcción de nuevas significaciones.

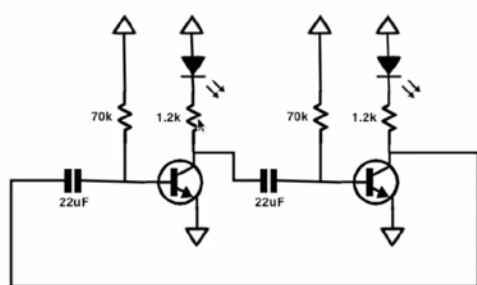


Imagen Bret Victor. Licencia CC BY-SA 4.0

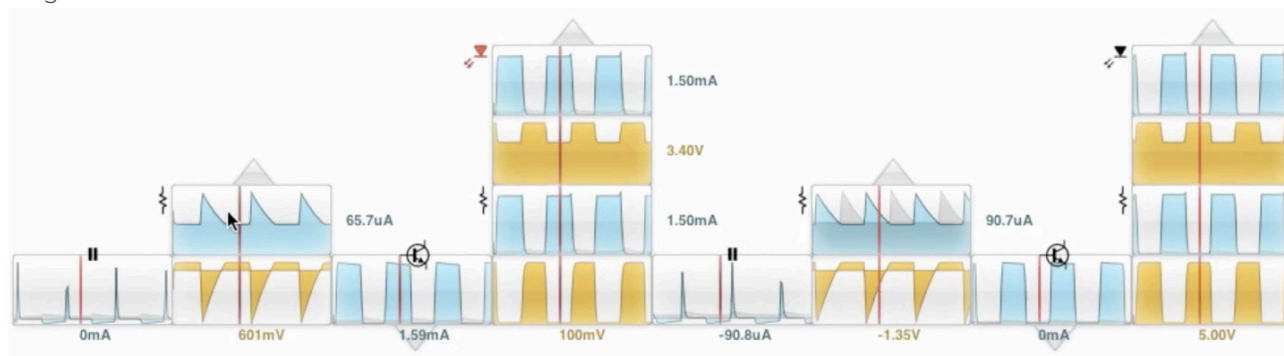


Imagen Bret Victor. Licencia CC BY-SA 4.0

En el prototipo interactivo que desarrolló, representaba el voltaje en color naranja y la corriente en color azul, además, cada símbolo electrónico era un elemento con el que poder interactuar y cambiar su valor, pudiendo ver, en tiempo real, cómo esta modificación afectaba a la dinámica de comportamiento global del sistema.

En este prototipo, el diseño de interacción era el elemento que nos permitía explorar otras formas de representación y, sobre todo, facilitar una mejor comprensión de este sistema eléctrico. Como os decía, el cambio de un parámetro (símbolo electrónico) nos permitía ver las implicaciones que tenía esta modificación en el resto del sistema.

Bret Victor propone la idea de entender los datos y la computación como un nuevo medio de

Como veis en la imagen, esta ha sido la representación clásica de un circuito eléctrico, donde podéis observar diferentes símbolos, que describen diferentes conceptos eléctricos, como la resistencia, el condensador, la conductividad, etc.

Esta representación nos ha permitido describir, y entender, el funcionamiento de un determinado sistema eléctrico. Bret Victor planteó si podríamos rediseñar este sistema de representación, de forma que nos permitiese tener una comprensión más profunda sobre la dinámica de comportamiento de estos sistemas, es decir que otras herramientas, podemos diseñar que nos permitan pensar, de otras formas, sobre la electricidad.

Bret Victor desarrolló este prototipo en el que exploraba qué nuevos medios computacionales, que interfaces, nos permitirían tener una mayor comprensión sobre los diferentes parámetros, conceptos electrónicos que influyen en el funcionamiento de estos sistemas eléctricos. Diseñó esta solución en la que proponía una interfaz en la que poder representar todos los parámetros que intervienen en el sistema.

pensamiento –*Computation as a new Media of Thought*-. Cómo el diseño de nuevas interfaces, sistemas de representación, y niveles de interacción, nos permiten describir una determinada realidad de diferentes formas, posibilitando una comprensión extendida y aumentada.

Necesitamos, por tanto, nuevos productos que posibiliten nuevas formas de relación con los datos y los algoritmos.

En la actualidad, estamos representando realidades muy diferentes con los mismos sistemas de información y esto no hace más que evidenciar los límites de la representación de información. Cómo, un mismo sistema de visualización, por ejemplo, un gráfico de barras, puede describir realidades tan diferentes como el promedio

de temperatura en un determinado año y país, o representar el número de ataques terroristas en Iraq. Esto lo que explicita, es la necesidad de construir nuevas narrativas en torno a los datos, explorar nuevas formas de visualización más empáticas que nos permitan humanizar y comprender mejor el comportamiento de los datos.

Paso 4.1. Un problema global.

El primer paso en el proceso de creación del proyecto será pensar en una determinada problemática global. Veremos cómo el diseño de otras inteligencias artificiales, puede ayudarnos a generar una nueva relación con el problema, extendiendo nuestro conocimiento.

El caso de uso que hemos utilizado, para describir el proceso propuesto en la guía, ha sido en el ámbito de la energía, donde exploraremos cómo el diseño de otras interfaces, inteligencias

¿Qué nuevas interfaces, qué nuevos sistemas de representación, nos pueden permitir representar una realidad, una determinada problemática, de otras formas?

artificiales, nos permitirá generar una nueva relación con los datos energéticos, con el objetivo de poder modificar y mejorar nuestros patrones de consumo eléctrico global.

En el apartado de materiales necesarios para ejecutar el proyecto, podéis ver diferentes problemáticas, que tienen como objetivo, que os puedan inspirar en la identificación del problema global a resolver.

Paso 4.2. Datos.

Es el momento, ahora, de comenzar a pensar en datos:

¿Qué tipo de datos, creéis, pueden estar ya disponibles en el problema global que hemos definido? ¿qué otras fuentes de datos podríamos utilizar en el caso de uso, de la energía, que hemos seleccionado?

En el caso de uso definido, por ejemplo, tendríamos datos relativos al consumo eléctrico en diferentes viviendas en las que se lleva a cabo el análisis, entre ellos:

Datos de consumo eléctrico por m² de vivienda.

Datos de consumo eléctrico por código postal y barrio.

Datos de consumo eléctrico por hora.

Datos de consumo eléctrico por nº de personas en la vivienda.

Datos Meteorológicos.

Nº de personas que habitan en la vivienda.

...

¿Qué otros datos, relacionados con la energía, creéis podrían ser interesantes para definir el caso de uso?

Os dejo unos minutos para que penséis.

...

Ya podemos continuar...

En este segundo paso, como veíamos en los materiales necesarios, os propongo que hagáis el siguiente, pequeño, ejercicio:

Seleccionar una app. que utilicéis con mucha frecuencia en vuestro móvil y me gustaría que anotéis todos los tipos de datos que creéis, puede el servicio, estar recogiendo de vuestra interacción con la app.

Por ejemplo, en Spotify, cada vez que utilizamos la app, están registrando grandes cantidades de datos, entre ellos:

Hora a la que accedo a la aplicación.

Durante cuanto tiempo uso el servicio.

Qué tipo de música y artistas suelo escuchar.

Qué tipo de música escuchan mis amigos.

Canciones que me gustan y comparto en redes sociales.

Canciones que nos recomienda la app y que no me gustan.

Qué canciones dejo de escuchar antes de finalizar la canción.

...

El objetivo de este segundo paso ha sido pensar en términos de datos y cómo estos modelan la

experiencia, y el tipo de relación, que tenemos con los servicios digitales que usamos.

¿Definiste ya el problema global sobre el que desarrollarás tu proyecto?, ¿qué tipo de fuentes de datos crees que estarían ya disponibles para poder usar?, ¿qué otros conjuntos de datos podríamos utilizar para enriquecer el problema global que queremos resolver?

Paso 4.3. Algoritmos.

En el siguiente paso, en el proceso de creación del proyecto, comenzaremos a pensar, de forma más detallada, cómo funcionan los principales tipos de algoritmos de inteligencia artificial y cómo podremos utilizarlos.

Cada tipo de algoritmo ha sido descrito, principalmente, por tres métricas: qué es, cómo funciona y en qué casos de uso podrá ser utilizado.

4.3.1. Aprendizaje supervisado

Aprendizaje Supervisado



Imagen Fran Castillo. Licencia [CC BY-SA 4.0](#)

Qué es:

El aprendizaje supervisado es un tipo de algoritmo que nos permite explorar diferentes tipos de problemas:

Problemas de clasificación binaria o multi-clasificación. En este tipo de problemas podremos predecir si un elemento pertenece o no a una

Qué es

El aprendizaje supervisado es un tipo de algoritmo que nos permite explorar diferentes tipos de problemas :

Problemas de clasificación binaria o multi clasificación (Variable objetivo = categoría) y **Problemas de regresión** (Variable objetivo = numérico)

Tipos de algoritmos

Linear Regression. Regresión.
Logistic Regression. Clasificación.
Decision Tree. Clasificación o Regresión.
Naive Bayes. Clasificación.
Support Vector Machine. Clasificación o Regresión.
Random Forest. Clasificación o Regresión.
AdaBoost. Clasificación o Regresión.
Gradient-boosting trees. Clasificación o Regresión.

Cómo funciona

1. Un humano etiqueta el conjunto de datos de entrenamiento y define la Variable objetivo.
2. El algoritmo es entrenado con un conjunto de datos históricos. Durante este proceso, extrae relaciones entre el input y el output (Variable objetivo).
3. Cuando el algoritmo es lo suficientemente preciso, podremos utilizarlo con un nuevo conjunto de datos.

Problemas de regresión. Este tipo de problema nos permitirá predecir, por ejemplo, el número de asistentes a una determinada actividad deportiva o la evolución de un precio de un producto en el mercado. La variable objetivo es un número.

Cómo funciona:

- I. Un humano etiqueta el conjunto de datos de entrenamiento y define la variable objetivo.
- II. El algoritmo es entrenado con un conjunto de datos históricos. Durante este proceso, extrae relaciones entre el input y el output (Variable objetivo).

III. Cuando el algoritmo es lo suficientemente preciso, podremos utilizarlo con un nuevo conjunto de datos.

Casos de uso Aprendizaje supervisado:

- Predicción con la probabilidad de abandono de un cliente en una compañía, por ejemplo, de telefonía.
- Predicción de ventas.
- Identificar cuando, en una tienda online, hay un cliente de alto valor, para poder personalizar la experiencia de cliente.

4.3.2. Aprendizaje no supervisado

Aprendizaje No Supervisado

Sin Variable Objetivo.



Imagen Fran Castillo. Licencia [CC BY-SA 4.0](#)

Qué es:

El aprendizaje no supervisado es un tipo de algoritmo que nos permite explorar diferentes tipos de problemas:

Problemas de clustering.

Perfilado.

Reducción de dimensionalidad.

Qué es

El aprendizaje no supervisado es un tipo de algoritmo que nos permite explorar diferentes tipos de problemas :

Problemas de clustering.

Perfilado.

Reducción de dimensionalidad.

Tipos de algoritmos

K-means clustering. Segment customers into group to better design mkt campaign or prevent churn.

Gaussian mixture model.

Hierarchical clustering.

Recommender system.

PCA o T-SNE

Cómo funciona

1. El algoritmo explora los datos de entrenamiento sin la necesidad de identificar una variable objetivo.

2. El algoritmo identificará grupos de datos con patrones de comportamiento similares.

Cómo funciona:

- I. El algoritmo explora los datos de entrenamiento sin la necesidad de identificar, previamente, una variable objetivo.
- II. El algoritmo identificará grupos de datos con patrones de comportamiento similares.

Casos de uso Aprendizaje no supervisado:

- Segmentación de clientes similares para optimizar el envío de campañas de marketing.

4.3.3. Aprendizaje por refuerzo.

Aprendizaje por refuerzo



Imagen Fran Castillo. Licencia [CC BY-SA 4.0](#)

Qué es:

Un algoritmo aprende a realizar una tarea simplemente tratando de maximizar las recompensas que recibe por sus acciones. Se produce un aprendizaje durante el proceso prueba/error.

Cómo funciona:

Cuando no disponemos de muchos datos de entrenamiento y no podemos definir el estado final ideal (variable objetivo), la única forma de aprendizaje es la interacción con el entorno.

El algoritmo realiza una acción sobre el entorno y recibe una determinada recompensa. El algorit-

Qué es

Un algoritmo aprende a realizar una tarea simplemente tratando de maximizar las recompensas que recibe por sus acciones.

Aprendizaje durante el proceso prueba/error.

Casos de uso

Optimizar la conducción en coches autónomos.
Optimizar el fijado de precios en tiempo real en un e-commerce, en función del volumen de interacciones.

Cómo funciona

Cuando no disponemos de muchos datos de entrenamiento y no podemos definir el estado final ideal (variable objetivo), la única forma de aprendizaje es la interacción con el entorno.

El algoritmo realiza una acción sobre el entorno y recibe una determinada recompensa. El algoritmo se optimizará a lo largo del tiempo para maximizar las recompensas.

mo se optimizará a lo largo del tiempo para maximizar las recompensas.

Casos de uso Aprendizaje no supervisado:

- Optimizar la conducción en coches autónomos.
- Optimizar el fijado de precios en tiempo real en un e-commerce, en función del volumen de interacciones.

Entendamos, ahora, cuáles son los principales algoritmos de redes neuronales, otra tipología de algoritmos de inteligencia artificial.

4.3.4. Redes neuronales convolucionales

Redes neuronales Convolucionales

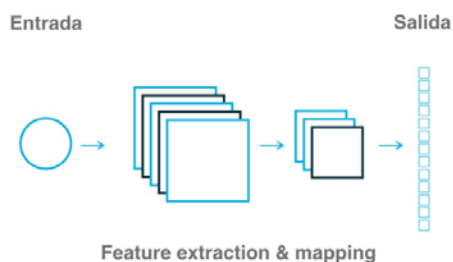


Imagen Fran Castillo. Licencia [CC BY-SA 4.0](#)

Qué es

Es una red neuronal multicapa con una arquitectura diseñada para extraer features de los datos de entrada.

Utilizaremos redes convolucionales cuando dispongamos de bases de datos no estructuradas. Son ideales para el procesamiento de imágenes y vídeo.

La principal diferencia de la red neuronal convolucional con el perceptrón multicapa viene en que cada neurona no se une con todas y cada una de las capas siguientes sino que solo con un subgrupo de ellas (se especializa), con esto se consigue reducir el número de neuronas necesarias y la complejidad computacional necesaria para su ejecución.

Cómo funciona

La red neuronal convolucional recibe, como input, una imagen y es procesada como una colección de píxeles (por ejemplo, letra L).

En las capas ocultas, se identificarán las características que permitirán, en este caso, detectar las líneas que forma una letra L.

La red convolucional clasificará una letra particular como L, si la red encuentra las características previamente identificadas en el proceso de entrenamiento.

Casos de uso

Diagnosticar enfermedades a través de imágenes, escáneres médicos. Identificación de objetos en una imagen.

Qué es:

Es una red neuronal multicapa con una arquitectura diseñada para extraer *features* de los datos de entrada.

Utilizaremos redes convolucionales cuando dispongamos de bases de datos no estructuradas. Las redes neuronales convolucionales son ideales para el procesamiento de imágenes y video.

La principal diferencia de la red neuronal convolucional con el perceptrón multicapa viene en que cada neurona no se une con todas y cada una de las capas siguientes, sino que solo con un subgrupo de ellas (se especializa), con esto se consigue reducir el número de neuronas necesarias y la complejidad computacional necesaria para su ejecución.

Cómo funciona:

La red neuronal convolucional recibe, como input, una imagen y es procesada como una colección de píxeles (por ejemplo, letra L).

En las capas ocultas, se identificarán las características que permitirán, en este caso, detectar las líneas que forman una letra L.

La red convolucional clasificará una letra particular como L, si la red encuentra las características previamente identificadas en el proceso de entrenamiento.

Casos de uso Aprendizaje no supervisado:

- Diagnosticar enfermedades a través de imágenes, escáneres médicos.
- Identificación de objetos en una imagen.

4.3.5. Redes neuronales recurrentes.

Redes neuronales Recurrentes

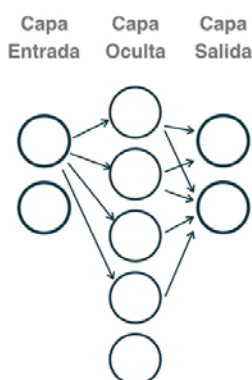


Imagen Fran Castillo. Licencia CC BY-SA 4.0

Qué es:

Las redes neuronales recurrentes no tienen una estructura de capas, sino que permiten conexiones arbitrarias entre las neuronas, incluso pudiendo crear ciclos, con esto se consigue crear la temporalidad, permitiendo que la red tenga memoria a corto plazo.

Cómo funciona:

Las redes neuronales recurrentes son muy potentes para todo lo que tiene que ver con el análisis secuencias, como puede ser el análisis y procesamiento de lenguaje natural (de texto y voz).

Qué es

Las redes neuronales recurrentes **no tienen una estructura de capas, sino que permiten conexiones arbitrarias entre las neuronas, incluso pudiendo crear ciclos, con esto se consigue crear la temporalidad, permitiendo que la red tenga memoria corta.**

Cómo funciona

Las redes neuronales recurrentes son muy potentes para todo lo que tiene que ver con el análisis secuencias, **como puede ser el análisis y procesamiento de lenguaje natural (de texto y voz.)**

Casos de uso

Chatbots.
Análisis del sentimiento, analizando, por ejemplo, datos de redes sociales.

Casos de uso Aprendizaje no supervisado:

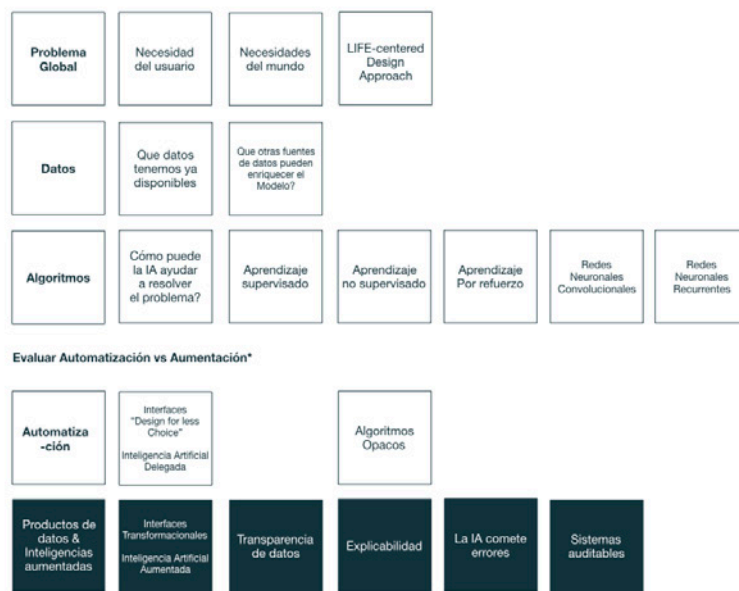
- Chatbots.
- Análisis del sentimiento, analizando, por ejemplo, datos de redes sociales.

¿Podríamos utilizar algunos de estos algoritmos para ayudarnos a resolver el problema global que hemos definido? ¿Cuál creéis que puede ser más relevante en vuestro proyecto?

A continuación, os presentaremos diferentes componentes, de la caja de herramientas, (Transparencia, Explicabilidad, Error, etc.) que os van a permitir poder explorar otras dimensiones de diseño en la construcción de inteligencias artificiales.

Paso 4.4. Una herramienta para diseñar inteligencias artificiales abiertas.

Una herramienta para diseñar inteligencias artificiales más humanas.



*Interfaces que extiendan nuestras capacidades y faciliten una nueva relación con el mundo.

Imagen Fran Castillo. Licencia [CC BY-SA 4.0](#)

A continuación, describiremos diferentes componentes que os puedan inspirar en el desarrollo de los proyectos.

El proyecto a desarrollar, podrá seleccionar varios de los componentes propuestos a continuación o, incluso, la totalidad de ellos. Cada componente,

debe ser entendido como un escenario desde el que inspirar el desarrollo de nuestro proyecto de inteligencia artificial.

Veamos, ahora sí, los diferentes componentes de diseño:

Paso 4.4.1. Nuevos patrones de diseño para la transparencia de datos.

El diseño para la transparencia de datos es ser totalmente transparente al usuario sobre los datos que la compañía dispone de nosotros y cómo van a hacer uso de ellos.

¿Qué nuevos patrones de diseño idearemos para facilitar una mayor transparencia de datos

en el diseño de productos digitales? ¿Por qué en la actualidad se están imponiendo los Patrones oscuros (Dark Patterns) en lo relativo al diseño con datos?



I Agree. Dima Yarovsky.
Licencia [CC BY-SA 4.0](#)



I Agree es una instalación artística que pretende problematizar cómo las principales compañías tecnológicas están haciendo frente a la privacidad de datos. La instalación es una visualización, una impresión del documento Términos & Condiciones que aceptas para poder hacer uso de un determinado servicio digital en Internet. La instalación artística representa el número de caracteres que dan forma al documento Términos & Condiciones de las principales aplicaciones: Facebook, Instagram, Snapchat, Tinder, etc.

Los datos son clave en la evolución de la Inteligencia artificial y son un activo diferencial para cualquier compañía en la actualidad. Hasta ahora, los datos han sido la moneda de intercambio por el uso de un determinado servicio en internet pero ahora podemos decir que este contrato social ya no es suficiente. Necesitamos que las compañías tecnológicas implementen, en sus productos digitales, nuevos patrones de diseño que nos permitan una mayor transparencia sobre los datos que disponen de nosotros y cómo van a hacer uso de ellos. En un escenario

de Grandes Datos, los usuarios tendrán el control total de sus datos (personales y de comportamiento) y podrán modificar el contrato cuando crean necesario.

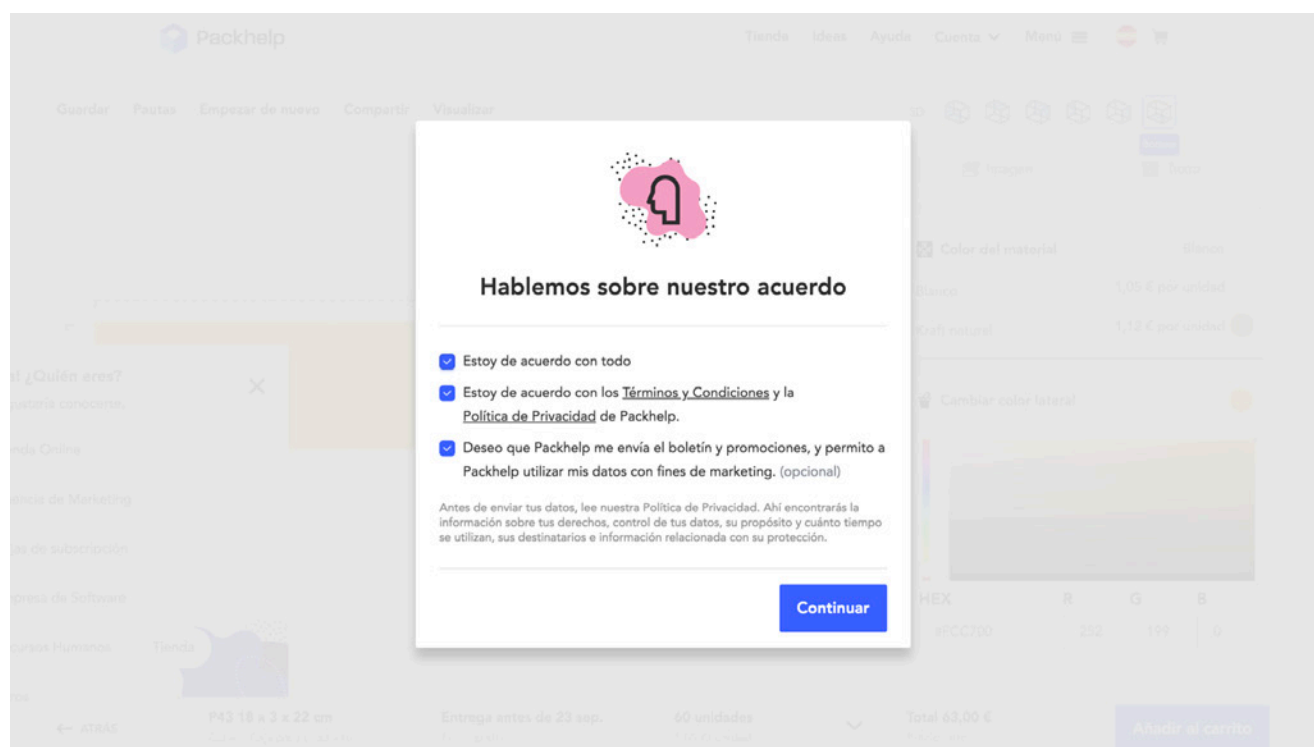
Cada vez somos más conscientes de la importancia de proteger la privacidad y exigir a estas compañías tecnológicas otro tipo de relación con nuestros propios datos.

¿Sabemos qué datos estamos compartiendo cuando usamos estos servicios digitales?

¿Sabemos cómo estas compañías tecnológicas están usando nuestros datos?

¿Confías en las compañías con las que estás compartiendo tus datos?

La transparencia parece ser, por primera vez, más valorada que nunca. Estudios recientes demuestran que el 95% de los clientes encuestados, comienzan a solicitar más información sobre cómo sus datos están siendo usados por estas compañías.



Término y Condiciones Packhelp. Licencia [CC BY-SA 4.0](#)

La industria digital y el diseño de producto han impuesto, como norma, el uso de *Dark Patterns* en todo lo relativo al diseño del tratamiento de los datos. Un *checkbox* junto a un texto de "Acepto y estoy de acuerdo con todo..." es la forma en la que, en la actualidad, estamos resolviendo esta problemática, que claramente no es suficiente y nos exige pensar en nuevas formas de diseño

que permitan al usuario el control, en todo momento, de sus datos.

Necesitamos nuevos patrones de diseño para la confianza.

Necesitamos una transparencia radical.

Necesitamos confiar en el ecosistema de datos de las compañías.

La agencia digital *GreaterThanExperience* está desarrollando un trabajo muy relevante en anticipar cuáles van a ser estos nuevos patrones de diseño en la construcción de soluciones digitales para una mayor transparencia de datos.

En cualquier producto digital:

Tú eres el propietario de tus datos.

Tú deberías poder tener el control de ellos.

Tú deberías obtener el principal beneficio de los datos que generas.

Quiero compartir un producto desarrollado por la agencia *GreaterThanExperience*. Es una solución que desarrollaron para la industria bancaria y en la que explicitan nuevos patrones de diseño que nos van a permitir una relación de mayor transparencia con los datos que generamos.

La interfaz va a solicitar de forma explícita los datos que necesita para proveer una determinada funcionalidad, permitiéndonos comprender, por qué necesita esos datos y como va hacer uso de ellos.

Imaginad, por ejemplo, que un determinado servicio digital va a implementar, en su aplicación, una funcionalidad para proveer ofertas personalizadas basadas en la localización.

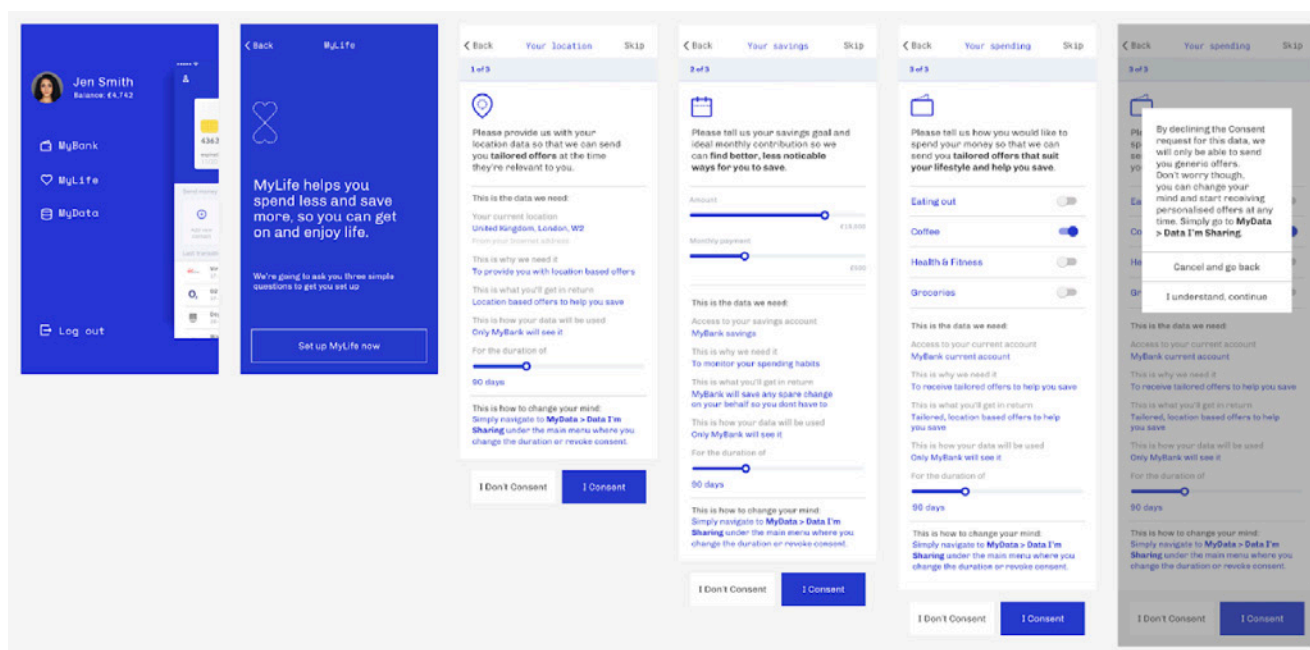


Imagen Agencia–*GreaterThanExperience*. Licencia [CC BY-SA 4.0](#)

La interfaz solicitaría la siguiente información:

- Estos son los datos que vamos a necesitar:
- Tu localización actual.
- Porqué necesitamos estos datos:

Para poder facilitarte ofertas personalizadas basadas en tu localización.

- Lo que obtendrás a cambio de poder compartir estos datos:

Recibirás ofertas que te ayudarán a mejorar tu salud financiera.

- Cómo vamos a hacer uso de tus datos:

Solamente el banco tendrá acceso a tus datos, no serán anonimizados ni compartidos con terceros. (*Data minimisation*)

Utilizaremos estos datos exclusivamente para esta funcionalidad.

- Duración:

Indicar el número de días que quiero tener acceso a esta funcionalidad.

Como podéis observar, la aplicación nos permite explorar qué otros patrones de diseño pode-

mos utilizar para posibilitar una nueva relación con los datos que generamos. Si nosotros, como usuarios, deberíamos ser los principales beneficiados de los datos que compartimos y deberíamos tener el control de ellos en todo momento, la aplicación propone la funcionalidad, de fácil acceso, para poder declinar el consentimiento cuando creamos oportuno, dejando de estar registrados nuestros datos (personales y de comportamiento) en sus sistemas.

En los productos de datos que diseñemos debemos proporcionar al usuario un control global de sus datos. Debemos permitir al usuario el control, y la personalización, sobre qué tipo de datos está monitorizando el sistema de IA.

Paso 4.4.2. Explicabilidad.

¿Por qué deberían los productos digitales que diseñemos integrar algoritmos que sean más interpretables, comprensibles y transparentes?

¿Por qué lo relevante, en la actualidad, no debería de ser el output del modelo, sino entender por qué un algoritmo está generando un determinado output?

¿Por qué necesitamos interfaces que nos posibiliten otra relación de mayor transparencia con los algoritmos? ¿Necesitamos interfaces que nos posibiliten una mayor comprensión sobre nuestros patrones de comportamiento, facilitándonos una mejor y consciente toma de decisiones? o, ¿Esto no es relevante?

Mientras la mayoría de los expertos, en la industria del Big Data y la Ciencia de Datos, tienen como objetivo último la construcción de algoritmos que tengan la máxima precisión, al mismo tiempo, es común, que estos expertos ignoren por qué un algoritmo está generando un determinado output. Los resultados de un algoritmo de Inteligencia artificial son, con frecuencia, percibidos como cajas negras (*black boxes*).

En la actualidad, los usuarios de servicios digitales están recibiendo continuamente recomendaciones de algoritmos basados en sus patrones de comportamiento pero sin embargo, desconocen cómo se están generando estas recomendaciones. Estos algoritmos de *machine learning* no están proporcionando un razonamiento. Por supuesto, no va a ser necesario mostrar el funcionamiento del algoritmo al usuario pero sí será

¿Qué es la minimización de datos?

La minimización de datos es crítica en el diseño de estos nuevos sistemas Humano-IA. El concepto de minimización de datos hace referencia al uso de la información estrictamente necesaria para cumplir una funcionalidad, es decir, no agrega datos complementarios para usos derivados.

Identifica, exactamente, qué datos necesitas del usuario. Si necesitas recopilar nuevos datos, solicita permiso explícitamente y describe para qué los va a utilizar y durante cuánto tiempo necesita conservarlos.

importante transparentar qué datos están siendo usados y el cómo se está generando un determinado output.

La transición hacia un entorno de *Explainable Artificial Intelligence* se acelerará principalmente a través de dos factores:

El primero es la demanda creciente, por parte de los usuarios, **de la necesidad de transparencia en las decisiones que están tomando los algoritmos de inteligencia artificial.**

Y el segundo, **la necesidad de poder tener una mayor confianza en los algoritmos.**

Proyectos de analítica avanzada, como el de *Scoring*, son casos de uso muy extendidos en la industria de servicios financieros, seguros y telecomunicaciones. Este caso de uso utiliza algoritmos de *machine learning* para asignar, a cada cliente, un score basado en sus datos históricos, con el objetivo, entre otros, de poder determinar la probabilidad de que un determinado cliente abandone la compañía o haga efectivo el pago de un préstamo bancario concedido.

En estos casos de uso, el algoritmo es una caja negra de la que sólo extraen valor las compañías, impidiendo al usuario poder activar sus propios datos. Si evolucionamos hacia un escenario en el que la mayoría de interacciones entre compañías y clientes estarán mediadas por algoritmos, vamos a necesitar que las compañías diseñen nuevas interfaces que nos permitan una mayor confiabilidad en los algoritmos.

Quería presentaros un servicio digital que opera en este nuevo escenario de diseñar productos digitales que proveen mayor interpretabilidad de los datos y los algoritmos y que posibiliten una mayor información y conocimiento al cliente. Además, esta transparencia, por parte de las compañías tecnológicas, se traducirá en una mayor confianza de los usuarios.

Credit Karma es una aplicación digital que provee al usuario, de forma completamente transparente, su score bancario basado en datos de su historial. A través de su producto, provee al usuario de información sobre qué tipo de datos y parámetros afectan a su *credit score* y sobre todo, qué puede hacer para mejorarlos. En esta solución, la explicabilidad del algoritmo hace referencia a la forma en la que éste procesa los datos y la importancia que tienen cada uno de los parámetros en la generación del score bancario.

Estos son algunos de los parámetros que describe el algoritmo: Uso de la tarjeta de crédito, Porcentaje de pagos que ha realizado a tiempo, Marcas derogatorias, Edad promedio de sus cuentas abiertas, Total de cuentas abiertas y cerradas, Número de veces que solicitó crédito. Cada parámetro describe el impacto que tiene en la generación del score.

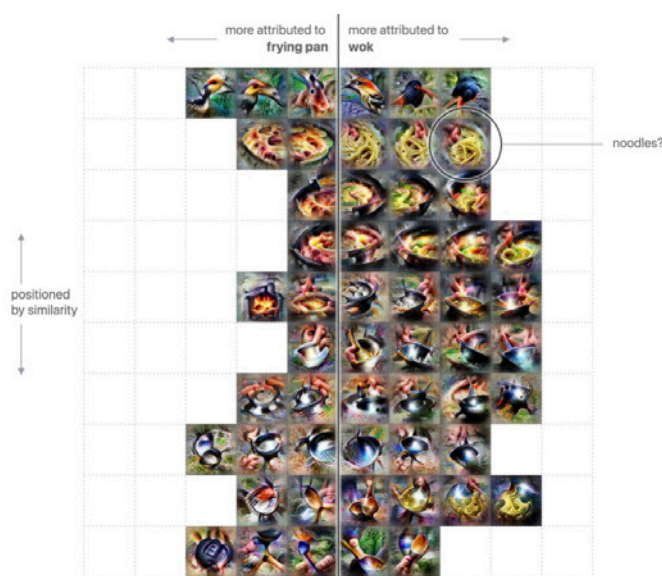
En la actualidad, los expertos en explicabilidad están inmersos en explorar el problema de los algoritmos de Inteligencia artificial como cajas negras. Este nuevo campo de investigación pone su foco en el diseño de IAs, más transparentes, que nos permitan entender mejor cuál ha sido el proceso de razonamiento seguido por el algoritmo.

A continuación, describiré diferentes técnicas de explicabilidad que serán muy útiles para entender como un algoritmo de inteligencia artificial generar un determinado resultado:

Activation atlases

Técnicas de Explicabilidad.

Activation atlases



Open AI and Google's activation atlas shows that this machine learning model has incorrectly learned to rely on the presence of noodles to distinguish woks from frying pans. (Open AI)

(Imagen Open AI & Google. Licencia [CC BY-SA 4.0](#))

Activation atlases es una tecnología desarrollada por Open AI & Google y cuyo objetivo es entender cómo un algoritmo de IA clasifica imágenes, además de poder, visualmente, identificar qué características tienen una mayor relevancia en la clasificación de una determinada imagen. Esta tecnología es, extremadamente, útil para identificar pequeños errores en los algoritmos de clasificación. En el ejemplo, podéis ver cómo el sistema de visualización desvela que el modelo, que fue diseñado para distinguir un *wok* de una

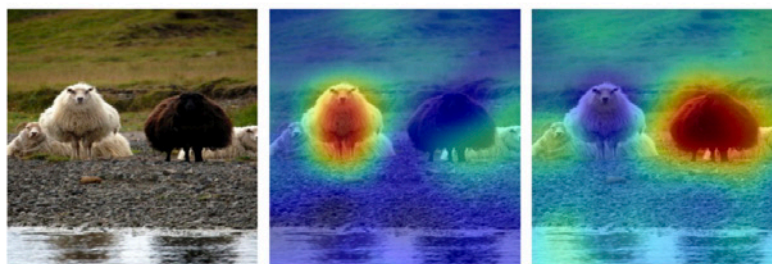
sartén plana, correlaciona de forma errónea, la presencia de *noodles* con un *wok*. La técnica de visualización lo que explicita es la necesidad de enriquecer el *dataset* para reconocer otras características, como la profundidad de *bowl* o el mango corto, en la identificación de un *wok*.

La visualización desvela la necesidad de reentrenar el modelo con otro tipo de imágenes que nos permitan captar mejor las características físicas de estos productos.

Saliency maps

Técnicas de Explicabilidad.

Saliency maps



(a) Sheep - 26%, Cow - 17% (b) Importance map of 'sheep' (c) Importance map of 'cow'

Saliency maps produced by the RISE method offer insights into which parts of an image an artificial neural network focuses on for classification. (Pascal Visual Object Classes)

Imagen *Saliency maps*. Licencia [CC BY-SA 4.0](#)

Saliency maps es otra técnica de visualización avanzada diseñada para conseguir que los sistemas de IA expliquen, o al menos parcialmente, sus decisiones, es decir, poder describir de forma más detallada como ha sido generado el *output* del algoritmo.

El sistema de visualización generará un mapa de calor que mostrará cuánto contribuyó cada uno de los píxeles en la clasificación de la imagen. En el ejemplo, se demuestra cómo un modelo, de

clasificación de imágenes, examinó una fotografía de una oveja blanca y otra negra y determinó que la oveja blanca era importante en la clasificación del objeto como una oveja, sin embargo el algoritmo determinó que la oveja negra probablemente era una vaca. Este resultado lo que desvela es la escasez de ovejas negras en el conjunto de datos de entrenamiento, para identificar correctamente la oveja negra.

Counterfactual explanations.

Técnicas de Explicabilidad.

Counterfactual explanations

Algorithm: "Your loan application was denied."

Applicant: "Why?"

Algorithm: "If you earned \$1,000 a month instead of \$750, it would have been accepted."

Applicant: "Disregarding my income and employment type, what can I do to get the loan?"

Algorithm: "You already have two loans. If you pay them back, you will get this loan."

Imagen <https://www.noemamag.com/ai-must-explain-itself/> Licencia [CC BY-SA 4.0](#)

Otro método utilizado, con mucha frecuencia, para explicar los outputs de los modelos de aprendizaje automático son los *counterfactual explanations*.

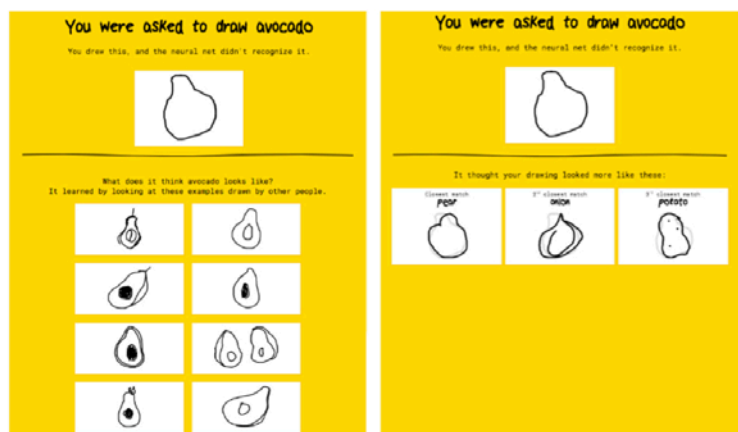
Esta técnica revela una comprensión basada en cómo los cambios mínimos en los datos de entrada generan un resultado diferente, lo que permi-

te comprender, de manera efectiva, cómo habría cambiado el resultado de un algoritmo si ciertas condiciones iniciales fueran diferentes. Por ejemplo, como veíamos, si una persona a la que se le niega un préstamo pudiese recibir una explicación sobre las variables que tienen una mayor influencia en la concesión del préstamo, esta información sería de gran valor para el usuario.

Explicaciones normativas vs explicaciones comparativas.

Técnicas de Explicabilidad.

Explicaciones Normativas vs Comparativas



Explicaciones normativas

Explicaciones comparativas

Imagen Quick Draw. Licencia [CC BY-SA 4.0](#)

Carrie Cai, investigadora de *Google Brain*, descubrió que diferentes tipos de explicaciones pueden llevar a los usuarios a diferentes conclusiones sobre la capacidad de los algoritmos. *Quick Draw* es una red neuronal artificial que tiene como objetivo identificar en tiempo real bocetos. Si la red neuronal no puede identificar el boceto, imaginad por ejemplo, un dibujo de un aguacate, el sistema ofrece dos tipos de explicaciones. Las explicaciones “normativas” muestran al usuario imágenes de bocetos de otros aguacates que el modelo identificó correctamente, para que el usuario pueda entender, de

forma muy visual, cuáles son los principales atributos en la identificación del boceto, mientras que las explicaciones “comparativas” se muestran, al usuario, cuando el modelo no reconoce el boceto, y este caso, muestra imágenes de bocetos con otras formas similares (como una pera o una patata).

Los usuarios que reciben explicaciones normativas tienden a confiar más en la inteligencia artificial, principalmente, porque provee mayor información que permite a los usuarios una mejor comprensión del sistema.

Paso 4.4.3. La Inteligencia Artificial comete errores.

Ayuda a los usuarios a comprender en qué escenarios el sistema de IA puede cometer un mayor número de errores.

Tradicionalmente hemos diseñado soluciones digitales que han tenido un *output*, *user flows* y experiencias de usuario predeterminadas. Sin embargo, diseñar con datos y algoritmos de *machine learning* es estar abierto a un cierto grado de indeterminación, imprecisión, de outputs o escenarios no predeterminados. Por tanto, diseñar con algoritmos de Inteligencia Artificial supone diseñar para muchos resultados diferentes.

Si los algoritmos de machine learning pueden, también, ser entendidos como un nuevo tipo de error (bug), ¿Qué nuevas experiencias de usuario

y patrones de diseño vamos a implementar para hacer frente a estos nuevos escenarios (errores) que pueden generar fricciones o incluso ser no agradables?

Nuestro sistema de IA no debería pretender ser siempre «inteligente», y proveer de una solución para todo, debe también, poder exponer sus propios límites de una manera transparente y honesta. No pretendamos que nuestro producto pueda tener una respuesta para todo.

A continuación, describo un ejemplo que evidencia la problemática de trabajar con algoritmos de aprendizaje automático en el diseño de servicios y productos digitales.

La Inteligencia Artificial comete errores.

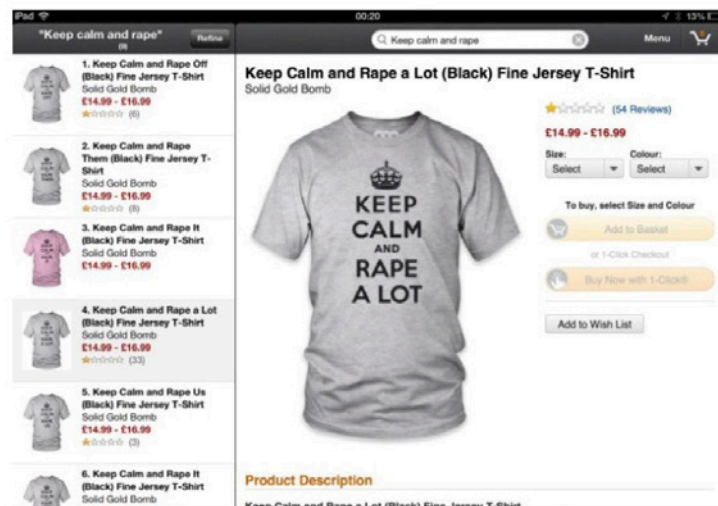


Imagen Amazon. Licencia CC BY-SA 4.0

Una compañía de impresión digital de camisetas decide construir un algoritmo para generar, de forma automática, multitud de combinaciones del popular mensaje motivacional, creado por el gobierno del Reino Unido, *"Keep Calm and Carry On"*. El algoritmo de generación de textos automáticos estaba conectado a la plataforma de Amazon, de forma que todas las variaciones del mensaje (*"Keep Calm and ..."*) eran publicadas en tiempo real. La denuncia por parte de los usuarios de Amazon que vieron la camiseta con este

mensaje impúdico, provocó que Amazon rescindiese el contrato, traducándose en pérdidas millonarias para la compañía de impresión digital. Lo que evidencia esta situación, es la incapacidad de las empresas para hacer frente a estos escenarios generados de forma algorítmica.

Necesitamos explorar nuevos patrones de diseño que puedan hacer frente, desde la disciplina de la experiencia de usuario (UX), a estos escenarios potencialmente desagradables.



Imagen Amazon. Licencia CC BY-SA 4.0

Otro ejemplo, muy interesante para entender cómo los algoritmos de IA cometen errores, es el trabajo de Joy Buolamwini. Joy desarrolla una investigación en la que pretende visualizar los sesgos existentes en los algoritmos de inteligencia artificial. Describe en su trabajo, cómo los principales software de reconocimiento de imágenes fueron incapaces de identificar los rasgos de su cara, simplemente por su color de piel, sin embargo, denunciaba de forma crítica, cuando se

ponía una máscara de color blanco, estos algoritmos sí que reconocían su cara, además de sus proporciones faciales.

Su investigación evidencia, y reclama, la necesidad de enriquecer, con mayor diversidad, las bases de datos que modelan, en la actualidad, los algoritmos. Necesitamos bases de datos en las que todas las singularidades puedan estar representadas.

Paso 4.4.4. Confianza y Algoritmos

En este último componente, quería compartir la necesidad de poder diseñar soluciones digitales, productos de datos, que nos posibiliten nuevas relaciones de mayor confiabilidad en los algoritmos.

Si evolucionamos hacia un escenario en el que las interacciones, entre marcas, compañías y usuarios, estarán mediadas a través del diseño de tecnologías de inteligencia artificial (chat-bots, asistentes digitales, etc) ¿Qué nuevos patrones de diseño, experiencias de usuario, van a posibilitar que podamos, como usuarios, tener una mayor confianza en los datos y en los algoritmos de estos servicios digitales? ¿Por qué deberíamos confiar en la recomendación de un determinado algoritmo?

En la actualidad, las compañías necesitan re-definir completamente la formas de relación con sus comunidades y construir nuevos servicios digitales que permitan otros tipos de relación

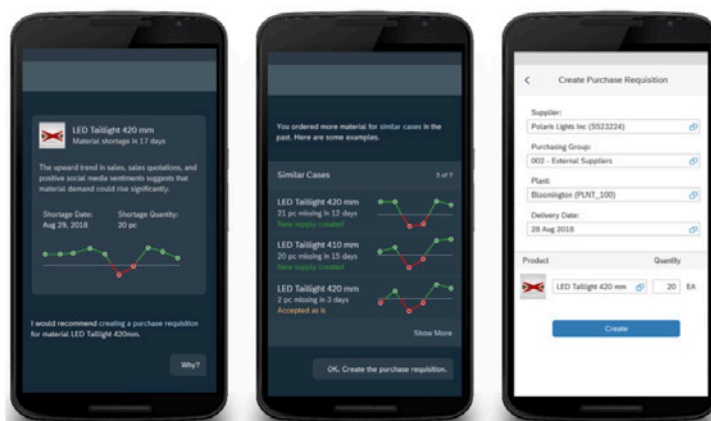
de mayor confiabilidad, posibilitando un mayor engagement. En este escenario, necesitamos idear nuevos patrones de diseño que nos permitan confiar en la manera en la que estos servicios y productos digitales están haciendo uso de nuestros datos pero también, entender la forma en la que estos algoritmos funcionan y nos están proveyendo recomendaciones.

Vladimir Shapiro propone tres dimensiones, que creo muy relevantes, para generar trust (confianza) desde el diseño y la experiencia de usuario (UX):

Describiré tres historias de usuarios para intentar entender cada una de las dimensiones propuestas y cómo incrementar nuestra confiabilidad en los algoritmos:

Demostrar competencia.

La Inteligencia Artificial debe demostrar competencia.



Vladimir Shapiro

Imagen Vladimir Shapiro. Licencia [CC BY-SA 4.0](#)

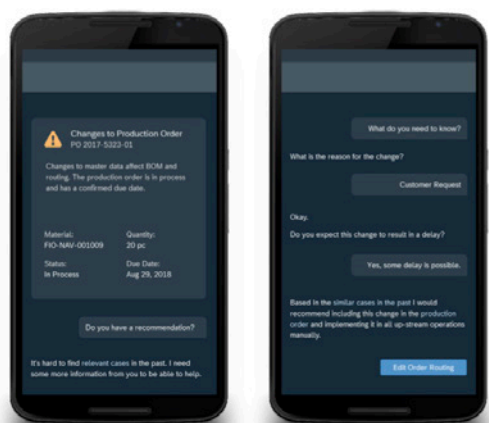
Paul es *retail manager* y está probando un nuevo asistente digital pero aún no confía mucho en él. Paul recibe una alerta sobre una nueva situación de rotura de stock de un determinado material. El asistente digital provee una recomendación basada en casos similares en el histórico de datos y describe su recomendación citando otros ejemplos similares en el pasado. Explorando otros casos similares en el pasado, Paul puede comprender la lógica detrás de la propuesta re-

comendada y puede confirmar, por su experiencia, que la recomendación es correcta.

Paul considera que su asistente es lo suficientemente competente para detectar este tipo de anomalías en los datos y ya está preparado para adoptar el próximo nivel de automatización.

Ser Auténtico y Honesto.

La Inteligencia Artificial debe ser auténtica & honesta



Vladimir Shapiro

Imagen Vladimir Shapiro. Licencia [CC BY-SA 4.0](#)

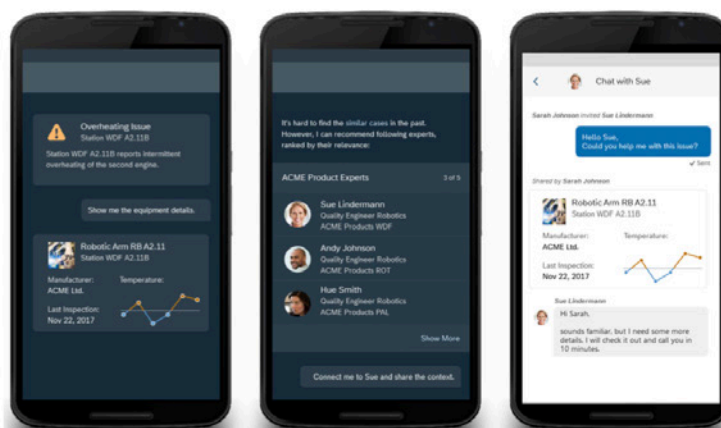
La segunda historia describe por qué un algoritmo o asistente digital debe ser honesto y mostrar también sus límites.

Como veíamos, Paul ya está familiarizado con el asistente digital y confía en sus recomendaciones, sobre todo en situaciones simples. En otros escenarios de mayor complejidad, Paul le pide recomendaciones a su asistente. Pero, ¿qué pasa si no hay suficiente información para que el asistente provea una recomendación de interés?

En lugar de pretender ser «inteligente», el asistente digital expone sus propios límites de una manera muy transparente, honesta y auténtica. El asistente hace preguntas apropiadas y finalmente juntos encuentran una solución.

Ser confiable.

La Inteligencia Artificial debe ser confiable.



Vladimir Shapiro

Imagen Vladimir Shapiro. Licencia [CC BY-SA 4.0](#)

Y para finalizar, la última historia de usuario.

Paul recibe una notificación sobre un problema en un proceso logístico. Su asistente no dispone de los datos suficientes como para ayudarlo con una recomendación o solución concreta. En este escenario, el asistente considera que es el momento de activar su modo *"I don't know. Wait a minute"* y lo conecta con un profesional experto.

Como veíamos en la historia de usuario anterior (*Being Authentic and Honest*), ser transparente con el usuario es una característica muy impor-

tante. En esta situación, el asistente no se rinde. En lugar de buscar otros problemas similares en el pasado, ajusta sus criterios y busca a profesionales expertos.

Por lo tanto, aunque el asistente digital no pudo proponer una solución específica, sí pudo conectar a Paul con otra persona para hacer un seguimiento correcto del problema detectado. En otras palabras, Paul puede confiar en su asistente digital para agotar todas las opciones disponibles que mejor se adapten a una posible solución.

Paso 4.4.5. Inteligencia artificial auditable.

Diseña un nivel de interacción que permita al usuario proveer de feedback sobre el resultado que genera el sistema IA.

La experiencia de usuario, de los productos que integran algoritmos de inteligencia artificial, mejorará en la medida en la que haya más datos en el sistema. Una mejor experiencia de usuario estará influenciada por el diseño de sistemas que nos permitan proveer feedback sobre el proceso de aprendizaje.

Describe cómo la interacción del usuario afecta al comportamiento futuro del sistema de IA. Ayuda a los usuarios a comprender cómo la forma en la que interactúan con el producto y/o algoritmo tendrá un impacto en el tipo de output que recibirán.

Imaginad algunas de estas aplicaciones:

La interfaz de un sistema de recomendación de libros debe de poder diseñar un nivel de interacción que permita al usuario, por cada libro recomendado, poder confirmar si esta información es, o no, relevante.

Cuando el usuario indique, por ejemplo, que el libro recomendado es de su interés, podría aparecer una información describiendo que recibirá otras recomendaciones similares.

Indique al usuario, con un mensaje o una descripción visual, que confirme que el sistema aprenderá del interés, o no, en el producto.

Permite al usuario ocultar un anuncio recomendado por el sistema y, al hacerlo, solicita feedback para poder mejorar la relevancia de futuros anuncios.

Informa al usuario cuando el sistema de IA actualice sus reglas de funcionamiento y describa cómo estas nuevas reglas influyen en el resultado proporcionado.

Evitar la percepción de desconfianza, por parte del usuario, cuando el algoritmo incorpora actualizaciones constantes sin notificar lo que ha cambiado.

Notifica al usuario cuando el algoritmo, por ejemplo, de ranqueo de contenido del feed de noticias cambie, describiendo qué variables influyen, ahora, en la información priorizada por el algoritmo.

5. Casos de uso.

5.1. Productos de datos e inteligencias aumentadas en un escenario energético y parkings inteligentes.

En este escenario de Grandes Datos (*Big Data*) y algoritmos de inteligencia artificial, es apropiado, como describíamos en la introducción, poder explorar nuevas visiones, desde el ámbito del diseño, que nos permitan idear nuevos productos digitales y que puedan ayudarnos a extender nuestra relación con los sistemas (ciudades, sistemas naturales, sociales, culturales, etc).

En este sentido, han surgido conceptos como el Diseño de Fricción (*Frictional Design*) o las Interfaces Transformacionales, interfaces orientadas a cambiar la forma en la que vemos e interactuamos con los sistemas.

E. Morozov en su libro “La locura del solucionismo tecnológico”, comienza el último capítulo –Dispositivos Inteligentes, Humanos Estúpidos– con esta cita de Bruno Latour:

“La ley moral está en nuestros corazones, pero también está en los artefactos (que diseñamos)”-Bruno Latour

Morozov quiere explorar con esta idea, el componente de no neutralidad en el diseño de tecnologías y productos que integran Datos y IA. Durante el capítulo, Morozov propone cómo el diseño de una determinada tecnología puede crear ciudadanos muy distintos.

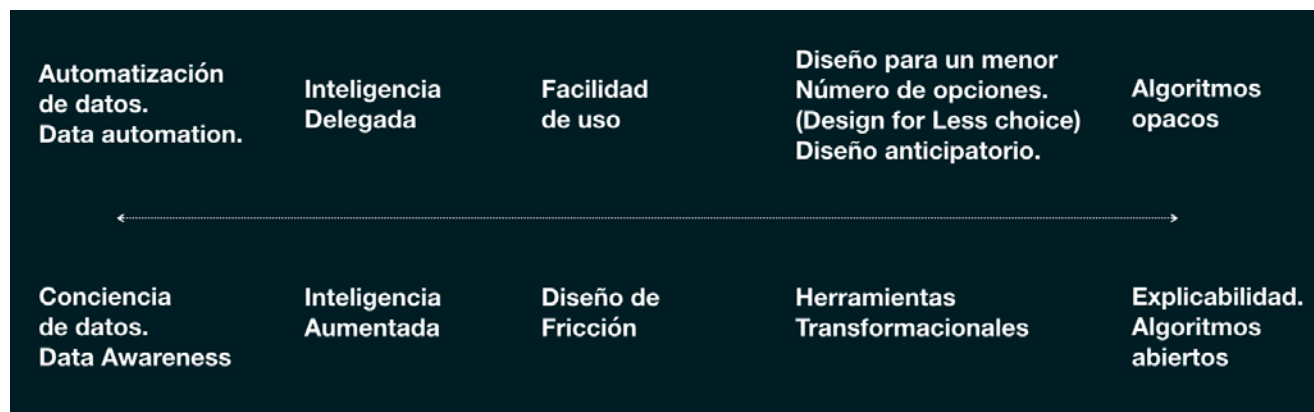


Foto Fran Castillo. Licencia [CC BY-SA 4.0](#)

Me gustaría explorar algunas de las ideas que expone Morozov y poder contraponer los dos paradigmas de diseño que os propongo, Diseño para un Menor Número de Opciones (*Design for less choice*) vs Interfaces Transformacionales. Utilizaré un caso de uso muy extendido en el contexto de las Ciudades Inteligentes (*Smart Cities*), en particular, una aplicación de Parking Inteligente (*Smart Parking*) diseñada para optimizar la gestión del tráfico.

El diseño de esta solución tecnológica propuso principalmente dos innovaciones:

La primera innovación, desarrolló un conjunto de sensores que permitirían que los parquíme-

tros se reiniciasen de forma automática cuando el coche abandonase el lugar en el que estaba estacionado, sin importar cuanto tiempo prepa- go quedaba aún en el medidor.

En la segunda innovación, el conductor no tendría manera alguna de sobrepasar el límite de tiempo fijado.

Estas dos innovaciones han construido una solución diseñada para maximizar la eficacia (económica) del sistema de gestión del parking en las ciudades, innovación que operaría en el primer principio – Diseño para un Menor Número de Opciones (*Design for Less Choice*). Soluciones y productos digitales que en su core limitan y con-

dicionan la forma de uso, de interacción con un determinado sistema (en este caso, el sistema de parking).

¿Qué sucedería si la tecnología, en lugar de reiniciar automáticamente el medidor cuando el conductor abandonase la plaza de aparcamiento, diera la opción al conductor de dejar el dinero excedente en el medidor, para que otro usuario pudiese hacer uso del tiempo restante?

Como describíamos anteriormente, E. Morozov explora cómo una misma solución tecnológica puede crear ciudadanos muy distintos. En este segundo escenario la tecnología no maximiza la eficacia económica del sistema, pero sí lo hace la deliberativa. Propone una nueva interfaz de relación que posibilita una mayor conciencia sobre el funcionamiento de este sistema.

La escalabilidad del primer principio es deficiente, porque para maximizar la eficacia de otros entornos, vamos a necesitar la implementación y despliegue de nuevos conjuntos de sensores. Sin embargo, el segundo escenario tiene una escalabilidad maravillosa, ciudadanos que construyen otra relación con las infraestructuras invisibles. En este caso, igual la recaudación de dinero no aumenta, pero sí lo hace la deliberativa, nuestra consciencia sobre el funcionamiento de los sistemas se incrementa.

Quiero compartir un proyecto que desarrollamos ya hace algún tiempo, y aún relevante para imaginar cómo pueden ser estas interfaces transformacionales que nos permitan entender mejor el funcionamiento de los sistemas, en este caso, como veíamos, en un escenario energético.

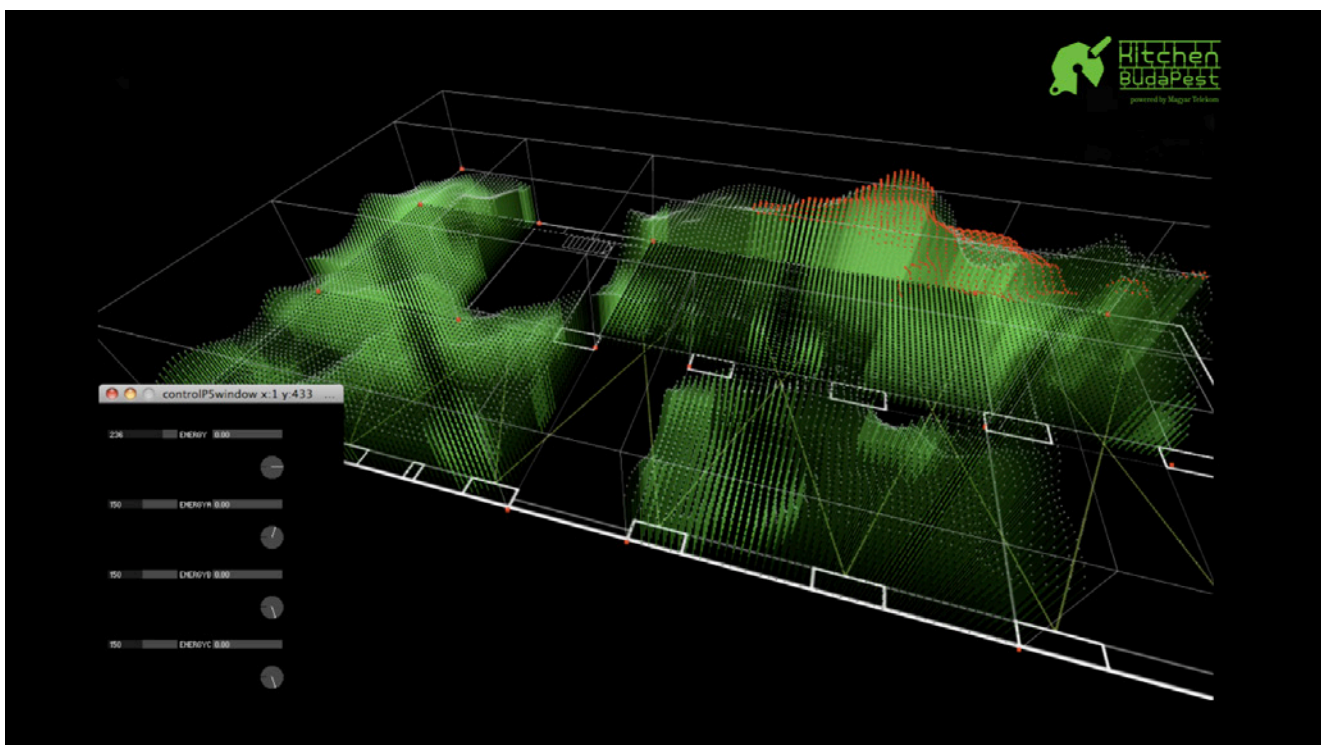


Foto Open Energy. Fran Castillo. Licencia [CC BY-SA 4.0](#)

Open Energy es un proyecto que tuve la oportunidad de desarrollar en Kitchen Budapest. Un Medialab financiado por Magyar Telekom. Desarrollamos una plataforma en la que poder monitorizar y explorar nuevas formas de visualización en tiempo real sobre el consumo energético en entornos domésticos y en las ciudades.

El proyecto planteaba la hipótesis, de cómo deberían ser estas nuevas herramientas transformacionales que posibilitasen una nueva relación con la energía. ¿Qué productos nos permitirían ser más conscientes de nuestros patrones de

consumo eléctrico? Y ¿Cómo podríamos mejorar nuestro comportamiento energético en casa?

Soluciones tecnológicas que apagan de forma automática un dispositivo cuando está en modo espera, son tecnologías de diseño que operan en el paradigma Diseño para un Menor Número de Opciones-Design for Less Choice (Automatización).

Pensemos por un momento, en el interruptor de la luz como una tecnología que nos permite un tipo de relación binaria con la energía (encendido- apagado). Este diseño nos permite un tipo de relación e interacción muy

simple con este sistema, impidiéndonos tener una mayor comprensión sobre nuestros patrones de consumo eléctrico.

Como proponíamos en el planteamiento del problema global, debemos poder ir más allá del paradigma de la eficacia y construir soluciones que mejoren la comprensión sobre nuestras dinámicas de comportamiento entorno al consumo eléctrico.

¿Cómo serían estas nuevas interfaces en el contexto de las infraestructuras de energía distribuida (*Smart Grids*)? Necesitamos de nuevas interfaces que nos permitan una nueva relación con la energía. Herramientas que posibilitan un mayor conocimiento sobre nuestros patrones de comportamiento eléctrico y sobre todo, herramientas que nos permitan cambiar nuestros hábitos.

Para finalizar, me gustaría que imagináseis ...

¿Cómo podría evolucionar la interfaz de Netflix en este nuevo paradigma del diseño transformacional?, ¿Cómo lo haría el sistema de recomendación de Amazon? y ¿la interfaz de Spotify?

A continuación, describiré otro caso de uso en el que hemos aplicado los diferentes componentes de diseño descritos en el paso a paso de esta guía.

En este caso, el problema global que hemos seleccionado, para el caso de uso, ha sido el COVID-19. El objetivo del proyecto ha sido explorar el desarrollo de nuevos, productos digitales y, patrones de diseño que nos permitiesen otras formas, más transparentes y confiables, de hacer un seguimiento de enfermos COVID-19.

¿Qué nuevos principios de diseño, creéis, deberíamos utilizar para desarrollar aplicaciones, bajo un nuevo enfoque, que contemplen una mayor transparencia de datos, explicabilidad y confiabilidad?

5.2. Principios de Diseño para APPS de rastreo de contactos afectados por COVID-19.

El objetivo de este caso de uso es presentar diferentes principios de diseño que deberían poder guiar la formulación de cualquier propuesta de aplicación digital de rastreo de contactos (*TACT-Technology-Assisted Contact-Tracing*) en un escenario de COVID-19.

La propuesta está articulada a través del paper de Daniel Kahn Gillmor – *Principles for Technology-Assisted Contact-Tracing*.

Cualquier aplicación digital de rastreo de contactos afectados por el virus del COVID-19, va a requerir de nuevos esquemas y patrones de diseño que nos permitan un tratamiento de los datos más transparente y confiable. Existen ya, diferentes propuestas de TACT que apuntan a estos nuevos patrones de diseño, como son, DP-3T, PACT, TCN y la propuesta de Apple y Google.

Antes de describir los diferentes principios de diseño, me gustaría compartir algunas cuestiones que creemos relevantes para pensar en el desarrollo de estas nuevas soluciones de rastreo:

¿Requieren estas aplicaciones de rastreo, un permiso explícito del usuario para poder instalarlas?

¿Puede una persona decidir no usar la aplicación sin ningún tipo de conflicto derivado?

¿Existen fuertes limitaciones del uso de los datos para fines distintos a los de salud pública?

¿Garantiza la tecnología, que los datos serán eliminados cuando no sean necesarios para el control de la epidemia del COVID-19?

¿Existen políticas y patrones de diseño establecidas para garantizar que solo se recopile la información exclusivamente necesaria? ¿Está prohibido de forma explícita compartir datos con entidades externas?

¿Es la tecnología transparente sobre los datos que se adquieren, cómo se usan y quién tiene acceso a ellos?

¿Los sistemas de almacenamiento de datos serán descentralizados o centralizados?

10 Principios básicos para el diseño de aplicaciones de rastreo de personas en el contexto de contención de la epidemia COVID-19:

01. Participación Voluntaria.

Cualquier uso de una aplicación digital de seguimiento de contactos, potencialmente afectados por el virus, requerirá de la voluntariedad del usuario en cualquier etapa del proceso de rastreo.

Cualquier patrón de diseño TACT, será efectivo en proporción a su adopción por parte de la ciudadanía, y esta adopción, dependerá en última instancia, de la transparencia y confianza sobre el tratamiento de los datos que haga la aplicación.

La aplicación tendrá que detallar, de forma explícita, qué datos personales serán registrados en el sistema, cómo se usarán y cómo serán protegidos.

El usuario decidirá de forma voluntaria:

Si instalar la aplicación de seguimiento en su teléfono.

Si llevar un teléfono con tecnología *Bluetooth*.

Si la App monitorizará su comportamiento en todo momento o podrá deshabilitarla durante los periodos, en los que el usuario no esté interactuando con otras personas, por ejemplo, cuando esté en casa o haciendo otra actividad aislada.

Si recibe una alerta debido a un contacto con una persona potencialmente expuesta a COVID-19, el usuario decidirá cómo reaccionar. Tendrá libertad para decidir selectivamente qué cantidad de registro de contactos (*contact log*) compartirá con la administración. Esto podría reducir el número de falsos positivos.

Si se diagnostica como infectado, el usuario decidirá si cargar su registro de contactos a la base de datos global.

El usuario podrá decidir, de forma voluntaria, si querrá que sus datos, anonimizados, formen parte de la base de datos para futuras investigaciones y análisis epidemiológicos.

02. Extender el Framework.

Una aplicación de seguimiento de contactos no podrá ser construida sólo por tecnólogos. Necesitamos extender el marco de trabajo e incorporar a profesionales de la salud pública y expertos en enfermedades infecciosas, de lo contrario, el proceso estará destinado a fallar.

Las tecnologías de detección de proximidad (*Bluetooth*, etc) no están capacitadas para evaluar, si la proximidad física detectada se traduce en un contacto epidemiológico. Por ejemplo, los dispositivos *Bluetooth* pueden indicar una proximidad entre individuos, pero sin embargo, éstos pueden estar en lados opuestos de una pared en la que un virus no puede cruzar.

¿Cómo vamos a determinar cuándo dos teléfonos han estado lo suficientemente cerca, el uno del otro, como para ser médicamente relevante?

¿Cómo vamos a evaluar esta proximidad en términos de contacto epidemiológico?

¿Qué datos deberían de activar una alerta por contacto?

¿La alerta se activará sólo si han estado cerca de alguien que ha experimentado síntomas de COVID-19 o sólo con personas que han recibido un resultado positivo acreditado con pruebas de laboratorio?

¿Cómo se priorizará la atención médica basada en el riesgo estimado detectado en la app?

La herramienta necesitará poder cambiar y adaptarse a cualquier circunstancia externa, así como tener los recursos para hacerlo de forma responsable.

La herramienta va a requerir de un marco de trabajo extendido con el soporte de diferentes expertos:

- Profesionales de la salud pública.
- Diseñadores UI / UX.
- Epidemiólogos.
- Criptógrafos.
- Investigadores de seguridad.
- Desarrolladores de software.
- Administradores de sistemas.

03. Preservando la Privacidad y la Transparencia de Datos.

Los sistemas digitales de rastreo estarán diseñados, exclusivamente, para realizar la función relacionada con la salud pública de contener la evolución de la epidemia. En ningún caso, recopilarán datos complementarios, que no sean estrictamente necesarios para su propósito de salud pública.

Si recordáis, en el componente sobre la transparencia, exploramos la importancia de incorporar nuevos patrones de diseño que faciliten una mayor transparencia de datos en la creación de productos digitales.

En este escenario de aplicaciones de rastreo:

Tus datos son tuyos.

Tu deberías de tener, en todo momento, el control de tus datos.

Tu deberías de obtener el principal beneficio de compartir tus datos.

Se deberán evitar el uso de datos que son controvertidos y de difícil anonimización (como los datos de localización). El diseño de aplicaciones, que preserven la privacidad, no debería de combinarse con sistemas que recopilan datos de difícil anonimización.

Cualquier identificador utilizado por el sistema no deberá estar vinculado a otros identificadores, incluidos no solo los números de teléfono y direcciones IP, sino también, otros identificadores anónimos.

El sistema obligará a mantener los datos encriptados. Una aplicación de seguimiento de contactos (TACT), no debe nunca almacenar datos de localización con su *timestamp* correspondiente. En un posible escenario de filtrado de datos, esta información sería potencialmente desfavorable.

04. Evitar los sesgos en los datos.

Hemos visto, durante los últimos años, sistemas tecnológicos de Inteligencia Artificial que afianzan, aún más, las desigualdades sociales. Diferentes ejemplos, como sistemas de reconocimiento facial, que tienen más probabilidades de identificar erróneamente a personas de color, o la asignación de créditos bancarios sesgados en función del sexo, etc.

El diseño de un sistema de rastreo de contactos asistido por algoritmos (TACT) deberá mitigar cualquier discriminación y desigualdad.

05. Tú tienes el control de los datos en cualquier lugar.

Las aplicaciones de seguimiento de contactos, con patrones de diseño *privacy-unfriendly*, están ya enviando una gran cantidad de datos a las autoridades centrales, dejando a los usuarios sin control sobre lo que sucede con estos datos

una vez salen de sus dispositivos.

Necesitamos nuevos patrones de diseño que faciliten un mayor control de nuestros datos. Estos patrones reducirán los riesgos derivados de la agregación de los datos por parte de la administración, como por ejemplo, que la administración decida utilizar esta información para controlar a las personas o incluso que esta información pueda verse técnicamente comprometida por terceros.

06. Minimización de datos en cualquier lugar.

¿Se está minimizando la recopilación de datos en las aplicaciones de rastreo?

¿Recopila la aplicación solo la información que estrictamente necesita?

Un patrón de diseño TACT que pretenda hacer frente a la totalidad de las pandemias futuras, necesariamente, requerirá recopilar información que no es necesaria para combatir la propagación del COVID-19. Este enfoque generalista, reduciría la confianza de los usuarios en el sistema, y la gestión de los datos adicionales desviaría los recursos de ingeniería necesarios para hacer la herramienta más efectiva y contrarrestar la actual epidemia.

La minimización de datos es crítica en el diseño de estos nuevos sistemas de rastreo de contactos afectados por el COVID-19, y debe de poder emplearse en cada paso del sistema de *trackeo*. El concepto de minimización de datos hace referencia al uso de la información estrictamente necesaria para cumplimentar una funcionalidad, es decir, no agrega datos complementarios para usos derivados.

¿Se destruirán los datos después de un período de tiempo?

Cualquier información almacenada en la base de datos deberá tener una fecha para la destrucción permanente de los datos, una vez que estos datos no sean epidemiológicamente relevantes.

La aplicación permitirá que el usuario pueda, de forma voluntaria, ceder sus datos anonimizados para futuras investigaciones y análisis epidemiológicos.

07. No Fuga de Datos.

Si los datos, en un patrón de diseño TACT, están disponibles por las autoridades centrales, y otros datos están disponibles en los dispositivos de los

usuarios, ninguna de estas partes debe divulgar deliberadamente datos a partes no involucradas. En particular, se deberán fijar reglas legales y técnicas para evitar que otras instituciones puedan acceder a cualquiera de estas bases de datos. Se combinarán con sistemas para detectar dichos accesos, así como, sanciones claras y exigibles por hacerlo.

08. Midiendo el Impacto.

Un sistema digital de rastreo de contactos debe facilitar a las personas una manera de saber si está funcionando o no. Estas métricas no necesitan ser perfectas o exactas, pero debe haber una idea aproximada de si el sistema, en su conjunto, está ayudando a reducir la transmisión del virus.

Esto podría ser tan simple como contar con un *dashboard* público que describiese diferentes indicadores como:

- El número de alertas móvil por coincidencia con afectados COVID-19.
- El número de descargas e instalaciones de la app.
- El número de contactos con el sistema médico a través de la app.
- El número de autoaislamientos.
- El número de afectados detectados de forma temprana con la aplicación.

Deberían de publicarse detalles sobre cómo se está midiendo el impacto, y cuáles son las métricas agregadas que nos permitan entender cuáles son los beneficios colectivos de la adopción de una aplicación de rastreo.

Consejos

Es importante recordar que cada uno de los pasos que os hemos propuesto para el desarrollo del proyecto deben de ser entendidos como componentes, de una caja de herramientas, que nos van a permitir enriquecer el diseño de nuestra solución, de Inteligencia artificial, en diferentes aspectos.

Cada componente ha presentado un escenario de reflexión sobre cómo pensar la evolución del producto de datos en un escenario de algoritmos de inteligencia artificial.

09. Definir una Estrategia de Salida.

Un esquema de privacidad TACT dirigido a contener la epidemia del COVID-19, no debería durar más allá de esta enfermedad en particular. Cualquier patrón de diseño TACT responsable, deberá de tener medidas incorporadas para su eliminación gradual.

Esto quiere decir:

- Saber hasta cuándo la aplicación ha sido útil e iniciar el cese del rastreo epidemiológico.
- Saber cuándo parar debido a la falta de impacto efectivo.
- Saber cómo apagar los servidores centrales de forma segura.
- Poder desinstalar la app correctamente en los dispositivos de todos los usuarios.

Y para finalizar con este principio, definir políticas de destrucción y borrado de datos, una vez haya finalizado el seguimiento de la población.

10. Un Sistema Auditable.

Y para finalizar, el décimo principio contempla que los usuarios y comunidades que se enfrentan con la oportunidad de participar en un proceso de rastreo de contactos, como medida de contención de la expansión de virus, necesitan indicadores claros de que el tipo de tecnología involucrado es confiable y hace lo que debe de hacer.

El sistema de diseño será completamente transparente para su revisión y mejora por cualquier usuario y/o organización. Se exigirá el uso de componentes desarrollados con tecnologías de código abierto.

Durante el desarrollo del proyecto, podrás utilizar la totalidad de los componentes presentados, o si lo prefieres podrás seleccionar, solamente, alguno de ellos, para integrarlos en vuestro proyecto.

El deseo de esta guía es que los diferentes conceptos, y pasos, presentados puedan ayudaros a entender el diseño de otras inteligencias artificiales.

Necesitamos, con urgencia, diseñar inteligencias artificiales, que no deleguen nuestra inteligencia en las máquinas sino, que extiendan nuestras capacidades cognitivas.

Como hemos visto, los datos y los algoritmos de IA ya están condicionando la forma en la que pensamos, vemos y nos relacionamos con el mundo. En este nuevo escenario, vamos a necesitar nuevas interfaces y productos digitales, que nos permitan que los algoritmos sean objetos observables, abiertos, y por lo tanto, objetos de discusión. Necesitamos, colectivamente, pensar en los efectos que imponen los algoritmos, pero sobre todo, herramientas que nos permitan una nueva relación con ellos.

Los tecnólogos y diseñadores deberíamos aceptar la idea de que nuestro objetivo no se limita a

lograr que las personas usen, de forma sencilla, los dispositivos, sino que además piensen gracias a ellos.

¿Te animas a imaginar cómo va a ser esta nueva generación de productos? :))

Nos encantaría poder conocer vuestras ideas, propuestas y entender en qué direcciones os ha inspirado esta guía.

Este es mi email: fran@datadrivenfutures.com, no dudéis en escribirme :)

Recursos

- Data-driven futures.
<https://datadrivenfutures.com/>
- UX en un escenario de inteligencia artificial.
<https://datadrivenfutures.com/2020/06/19/ux-y-ai/>
- Artificial intelligence must explain itself.
<https://www.noemamag.com/ai-must-explain-itself/>
- Principles for Technology-Assisted Contact-Tracing by Daniel Kahn Gillmor.
<https://www.aclu.org/report/aclu-white-paper-principles-technology-assisted-contact-tracing>
- La locura del solucionismo tecnológico. Evgeny Morozov.
- Open AI.
<https://openai.com/blog/introducing-activation-atlases/>
- Técnicas de visualización avanzada.
<https://distill.pub/>
<https://distill.pub/2019/activation-atlas/>
- Bret Victor.
<http://worrydream.com/>
- Ux for trust.
<https://medium.com/sap-design/ux-for-ai-trust-as-a-design-challenge-62044e22c4ec>
- Greater than experience. Ethics, Privacy and trust.
<https://www.greaterthanexperience.design/>
- Quick draw
<https://quickdraw.withgoogle.com/>

Ciencia Ciudadana

Las infraestructuras



la aventura
de aprender