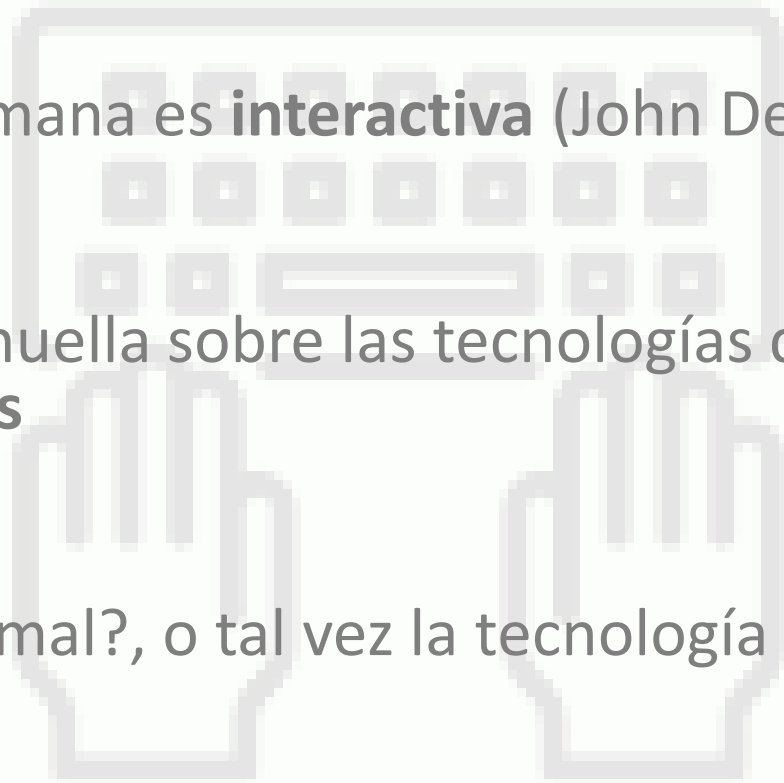




# ¿Inteligencia artificial o sistemas de control delegado?


# Interacción

- La experiencia humana es **interactiva** (John Dewey)
- Dejamos nuestra huella sobre las tecnologías digitales... y este proceso deja **huella en nosotros**
- ¿Para bien o para mal?, o tal vez la tecnología sea **neutral**



# ¿Neutralidad de la técnica?

- La técnica, en general, es **necesaria** para la vida humana. No es neutral, es buena
- En un segundo sentido, se puede decir que una herramienta técnica concreta, como un **cuchillo** es neutral, pero otras, como una **bomba** atómica parecen no serlo tanto
- Depende también la neutralidad del **nivel** que analicemos: la tecnología nuclear puede tener aplicaciones buenas y malas; no así, en concreto, la bomba atómica. La clonación, como biotecnología general, tiene aplicaciones buenas y malas; no así la clonación humana en concreto
- Cada tecnología cambia nuestras **capacidades**, pero también nuestras **necesidades**. Se pone a nuestro servicio al mismo tiempo que nos somete a servidumbres... y hay servidumbres inaceptables. En este sentido tampoco hay neutralidad
- **Elegimos** entre líneas de desarrollo tecnológico, unas mejores que otras



“Podemos dar el *sí* a la ineludible utilización de los objetos técnicos, y podemos a la vez decir *no* en cuanto les prohibimos que exclusivamente nos planteen exigencias, nos deformen, nos confundan y por último nos devasten”

Heidegger, *Serenidad*, 1949

# IA, ¿Sí o no?


- Depende de las **aplicaciones**, sí, pero estas dependen, a su vez, de la **interpretación** (ontológica, epistemológica y práctica) que hagamos de los llamados sistemas de IA
- Habría que decir **NO** a los desarrollos y aplicaciones de sistemas IA inspirados en una **interpretación incorrecta**, porque tienden a devastar la vida humana
- **SÍ** a los desarrollos y aplicaciones digitales basados en una **interpretación correcta**, porque sirven para mejorar la vida humana, para hacerla más propiamente humana



Fuente:

# Consideraciones ontológicas

- ¿Existe la **IA**?
  - Luc Julia, *L'Intelligence artificielle n'existe pas*, 2019
  - Marie David y Cédric Sauviat, *Intelligence Artificielle. La nouvelle barbarie*, 2019
  - Richard Benjamins e Idoia Salazar, *El mito del algoritmo*, 2020
  - Erik J. Larson, *The Myth of Artificial Intelligence*, 2021
    - “Inductive AI will continue to improve at narrow tasks, but if we want to make real progress, we will need to start by more fully appreciating the only true intelligence we know—our own”
- Existen los **sistemas de IA**. En ellos lo artificial no es inteligente, lo inteligente no es artificial
  - Hay que pasar de considerar los sistemas de IA como "sistemas técnicos con consecuencias sociales" a considerarlos como "sistemas sociales técnicamente implementados" R. Hirschheim, H. K. Klein y K. Lytinen, *Information Systems Development and Data Modelling*, CUP, 1995
- El **nombre** nos desorienta (introduce supuestos abusivos, orienta mal la interpretación de la cosa, su desarrollo y aplicaciones) ¿Otro mejor?
  - *Machine learning, deep learning...*
  - Inteligencia asistida, ampliada, centrada en humanos (*Stanford Institute for Human-Centered AI*, Fei-Fei Li y John Etchemendy), herramientas de apoyo a la decisión (Sara Lumbreras, UPCO), sistemas de control delegado (sistemas **CoDe**; *DeCo systems*)...



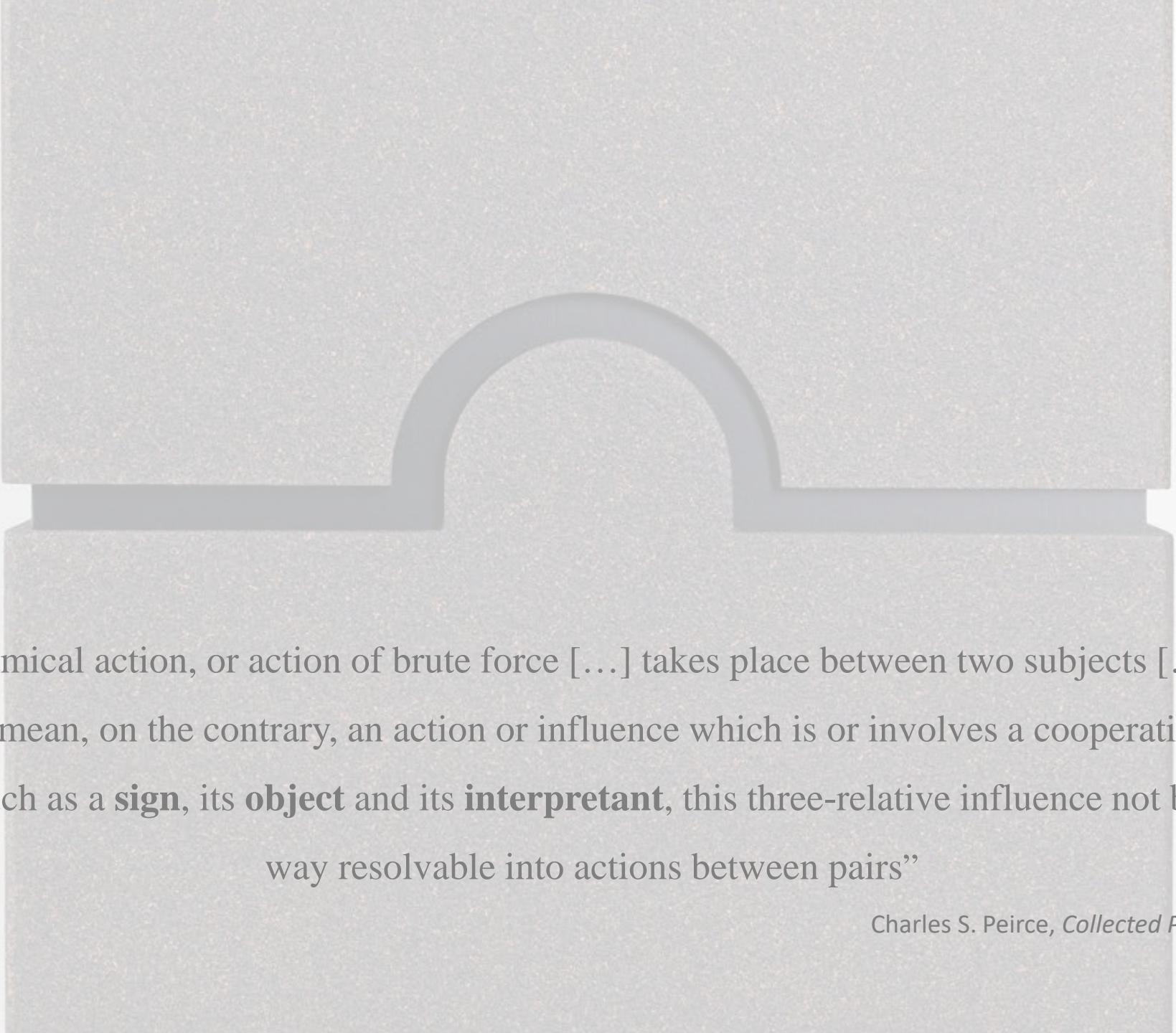
“**El término inteligencia artificial** surgió en los años cincuenta, cuando los científicos querían recaudar dinero para sus investigaciones. Pensaron que sonaba a algo que el Estado fomentaría de buen grado. Y ahora pendemos de este nombre. La mayoría de **los científicos informáticos lo encuentran inapropiado**”

Katharina Zweig, Laboratorio de Responsabilidad Algorítmica, U. de Kaiserslautern, 2021



# Consideraciones ontológicas


- ¿Qué es la inteligencia?
  - **Intelligere**, *intus/inter legere, legein* (λέγειν, hablar, decir, relatar, de ahí λόγος; en griego arcaico, de Homero, seleccionar, recolectar, enumerar), *leg-* (elegir-entre, diferenciar). Leer-dentro, ir a la esencia y las causas, ligar, unir lo semejante, reunir en el concepto y, en última instancia, en la unidad de la conciencia.
  - Capacidad de **entender** (pensar). Una máquina no entiende
  - Capacidad para **resolver problemas**. Los problemas son siempre del usuario humano, no de la máquina (solo los seres vivos tienen problemas)
  - Stuart Russell & Peter Norvig, *AI. A Modern Approach* (4th ed. 2021): thought vs. behaviour / human vs. rational
- Simulación de funciones de la inteligencia humana (*The Great Pretender, The Platters*, 1955)
  - **Simular no es ser** (el no ser  $x$  es condición de posibilidad para simular  $x$ )
  - Lo que hace la máquina solo cuenta como **simulación** y como **función** para un humano
  - **El punto de vista humano cambia la ontología de la máquina**



“All dynamical action, or action of brute force [...] takes place between two subjects [...] But by semiosis I mean, on the contrary, an action or influence which is or involves a cooperation of **three** subjects, such as a **sign**, its **object** and its **interpretant**, this three-relative influence not being in any way resolvable into actions between pairs”

Charles S. Peirce, *Collected Papers*, 1931-1935

# Consideraciones ontológicas

A green, textured dinosaur-like creature is shown from the chest up, holding a white computer keyboard in its mouth. The creature has large, expressive eyes and a slightly open mouth. The background is a blurred, dimly lit room with some lights visible.

- Sin *interpretant*, la máquina ya no es parte de un sistema de IA, es solo materia sometida a cambios físicos
  - *Mecanópolis* y efecto *Toy Story*
  - Necesidad de **man-tenimiento**
  - Sin mirada humana no hay **datos** (y menos aun **datos sintéticos**), ni **funciones**, ni **simulación**, ni **problemas**, ni **información**, ni **decisión**, ni **realidad virtual...** ni **IA**. No hay ni siquiera **computo** (Borges/Nietzsche)
- Seguir llamando inteligente a la parte artificial de un sistema...
  - **Devalúa** la inteligencia
  - Fomenta y supone una concepción **dualista** de lo humano

# Consideraciones epistemológicas: correlación e inducción

- Detección de correlaciones en *big data*
  - No son relaciones **causales**
  - No permiten ni **entender** ni **explicar**
  - Alto valor **heurístico**
- Los sistemas IA están afectados por los problemas de la inducción
  - Generan **expectativas** a partir del histórico de **datos**
  - Pueden ser **correctos los datos** y **fallar el sistema de expectativas**
  - **No predicen**, establecen modelos, no hablan en futuro, sino en **condicional**
  - La **condición crucial** es que todo siga siendo como viene siendo
  - **Regularidad y novedad** en el universo



“El porvenir está **abierto**”

karl Popper, 1995

“Nosotros sabemos —y tal vez es lo único que sabemos— que la mayoría de las cosas serán distintas [...], que hemos de contar siempre con la **novedad**, pero que no sabemos calcularla”

Hans Jonas, 1979

# Consideraciones epistemológicas: abducción

- Aprendemos de la **experiencia**, pero **sin garantía** de que las cosas sigan siendo como venían siendo (problema de la **inducción**)
- La **abducción** es un tipo de inferencia dinámica y falible que integra inducción, deducción y momento **creativo** (C. S. Peirce; A. Aliseda)
- Humildad intelectual, actitud socrática, prudencia aristotélica, de docta ignorancia, de **falibilismo**...

## Consideraciones epistemológicas: Una inteligencia prudencial en un universo abierto

- Solo una inteligencia viva, sentiente, situada, una **inteligencia prudencial** como la humana, puede entenderse con esta **textura del universo**
- El **“método” más abarcador**, el que regula la aplicación de todos los demás, incluidos los automatizables, es la prudencia, el sentido común
- La **prudencia** nos facilita la constitución **integradora de la experiencia**, la **gestión de las emociones** vinculadas a la frustración de expectativas, la **propedéutica del momento creativo** y el **filtrado crítico** de los sistemas de expectativas emergentes
- **Fuerza bruta de computación**, sí, pero también la imaginación, creatividad, intuición y... **prudencia**



“Código de conducta articulado en torno a seis puntos: 1. **Prudencia...**”

*Declaración de Barcelona para el desarrollo y uso adecuado de la IA en Europa, 2017*




# Una inteligencia prudencial en un universo abierto

- En consecuencia, un sistema de **IA** que pretenda **sustituir a la prudencia humana** simplemente estaría **fuera de lugar**, fuera del universo que nos alberga
- Un sistema **CoDe** inscrito en el **marco de una inteligencia humana prudencial** estará en el lugar que le corresponde y podrá cumplir funciones de gran valor para la vida humana

# Consideraciones ético-políticas

- Dadas sus características ontológicas y epistémicas, ¿cómo debemos manejar los sistemas CoDe?
  - La **decisión** (humana) de **delegar el control** (en máquinas) genera **responsabilidades** (humanas)
  - Para dirimir estas responsabilidades hay que **preguntarse** caso por caso:
    - ¿Quién delega?, ¿está legitimado para hacerlo? ¿En qué sistemas CoDe delega?, ¿son los más apropiados?, ¿son fiables, auditables, transparentes, explicables?, ¿es imprescindible que lo sean? ¿Qué tipo de acciones son delegadas?, ¿resultan, de verdad, delegables? ¿Por cuánto tiempo se cede el control?, ¿es sensato? ¿Es reversible la delegación? ¿Qué procedimientos de supervisión o evaluación existen?, ¿son suficientes? ¿Qué riesgos se asumen en caso de fallo?, ¿es prudente asumirlos? ¿Qué ventajas se obtienen para la vida humana con la delegación de control?, ¿qué se pierde a cambio?...



“Lo que debería aterrorizarnos no es un **futuro** dominado por una hipotética IA superior [...] Lo que realmente debería preocuparnos es la situación **presente**, en la que estamos **delegando** cada vez más tareas en una IA tan limitada como la actual”


Ramón López de Mántaras, fundador del Instituto de Investigación en IA del CSIC, 2020

# Consideraciones ético-políticas

- Educación práctica en **virtudes** (prudencia, creatividad, laboriosidad, desasimiento...) en el contexto tecnológico actual
- Consideraciones **políticas**: demasiado poder en pocas manos (PCCh-BATX, GAFAM)
  - ¿Es legítimo?, ¿la libertad amenazada?, ¿qué se puede hacer?, ¿uso responsable de lo digital por parte del ciudadano?, ¿denunciar el caso chino, fraccionar compañías americanas?, ¿presión ciudadana?, ¿equidistancia europea?
  - **Derechos digitales**: Ni los robots ni el software tienen derechos. Los humanos sí, pero no hace falta inventar nuevos derechos digitales, sino derivar la legislación de lo digital de los genuinos derechos humanos

# Resumen conclusivo

- Nuestra acción es siempre **interacción**
- Lo técnico **no** (siempre) es **neutral**
- ¿Cómo **valorar IA**? Primero **ontología** y **epistemología**, después **ética**
- Ontología: “La IA no existe”, existen los sistemas **CoDe**
- Epistemología: Los sistemas CoDe no predicen, son falibles, son vulnerables al **problema de la inducción**, dependen de la prudencia humana
- Consideraciones prácticas:
  - La **responsabilidad** es humana, hay que juzgar **caso por caso** la decisión de **delegar el control**
  - Educación en **virtudes** a la altura del actual contexto tecnológico
  - Es recomendable **romper la actual concentración de poder digital**
  - **No** lo es, en cambio, promover cartas de **derechos específicamente digitales**

A hand is shown holding a glowing blue brain that is connected to a circuit board. The background is dark blue with a network of white nodes and lines, some of which are glowing. The overall theme is artificial intelligence and human-machine interaction.

¡¡¡Muchas gracias!!!