

Crítica de la razón digital

El desafío de la humanidad digital

Ética y religión en tiempos digitales

ULRICH HEMEL



FONDO
EDITORIAL



Crítica de la razón digital

V O L U M E N 3

El desafío de la humanidad digital
Ética y religión en tiempos digitales

ULRICH HEMEL



Dedico este tomo a mis tres nietos, Justus (2013),
Jonas (2015) y Amalia (2019).

Hemel, Ulrich, 1956-

Crítica de la razón digital : volumen 3 : el desafío de la humanidad digital. Ética y religión en tiempos digitales / Ulrich Hemel -- 1a ed. -- Lima : Universidad San Ignacio de Loyola. Fondo Editorial, 2021.

3 v. : 21 cm.

ISBN: 978-612-4370-72-4 (obra completa)

ISBN: 978-612-4370-73-1 (v. 1)

ISBN: 978-612-4370-79-3 (v. 2)

ISBN: 978-612-4370-80-9 (v. 3)

1. Internet--Aspectos morales y éticos. 2. Internet -- Aspectos religiosos 3. Tecnología de la información -- Aspectos sociales. 4. Internet -- Aspectos sociales.

302.23

H38

CRÍTICA DE LA RAZÓN DIGITAL

Volumen 3:

El desafío de la humanidad digital.
Ética y religión en tiempos digitales.

© Ulrich Hemel

Primera edición, julio de 2021

© De esta edición

Universidad San Ignacio de Loyola
Fondo Editorial
Av. La Fontana 750, La Molina
Teléfono: 3171000, anexo 3705

Institut für Sozialstrategie
89150 Laichingen
Bleichwiese 3

kontakt@institut-fuer-sozialstrategie.org

Director: José Valdizán Ayala
Coordinadora: María Olivera Cano
Editor: Rafael Felices Taboada

Diseño y diagramación: Sergio
Pastor Segura

Traducción: LASA Servicios
Generales S.A.C.

Las imágenes fueron descargadas de
freepik.com y *unsplash.com*

Hecho el Depósito Legal en la
Biblioteca Nacional del Perú
Nº 2021-02521

Impresión bajo demanda
Aleph Impresiones S.R.L.
Jr. Risso 580 - Lince

Julio 2021

Tiraje 50 ejemplares

Se prohíbe la reproducción total o
parcial de este libro, por cualquier
medio, sin permiso expreso del
Fondo Editorial.

Contenido

1. Ética digital	12
El arte de la argumentación ética como tarea en el mundo digital	17
La humanidad en el contexto de la dignidad humana y los derechos humanos	22
La humanidad, los robots y las leyes de Asimov	25
Ética utilitaria, deontológica y de la virtud: la humanidad como criterio	32
Arquitectura mental y contextos situacionales de acción	42
El desafío de la responsabilidad: métodos de toma de decisiones éticas	50
Reflexión sobre principios: la responsabilidad digital en su sentido moral y legal	57
Ética en el mundo digital y ética profesional digital	68
La ética como parte de la tecnología digital: ética por diseño y diseño sensible a valores éticos	72
La equidad digital como tarea de empresas e instituciones	75
La estrella guía de la humanidad: valores humanos para el mundo digital	80
Referencias bibliográficas	84
2. Religión digital y humanidad digital	92
La imagen de Dios: la dignidad de las personas y las máquinas	97
Superinteligencia digital e inmortalidad digital	102
La religión digital como esperanza de salvación digital	108

Práctica religiosa en mundos digitales y teología digital	110
Creación digital: máquinas digitales como cuasi-personas	113
El reto de la humanidad digital	123
Referencias bibliográficas	127

Prefacio

¿Seguirán existiendo los seres humanos en la época digital? ¿Qué quedará cuando una superinteligencia regule nuestro comportamiento humano hasta el último detalle? ¿La creación digital hace que su creador, el ser humano, sea superfluo? ¿Y la ética sigue teniendo algún sentido cuando las decisiones son tomadas básicamente por programas inteligentes en los que no podemos influir?

Al final de nuestro gran viaje hacia el mundo digital, se plantea la cuestión de la humanidad. Una cosa está clara: no se puede responder a estos temas sin la pregunta sobre el sentido de nuestras vidas y los objetivos comunes. La ciencia por sí sola no puede dar esta respuesta: plantea hipótesis que luego se verifican o falsean. Pero no proporciona respuestas a las cuestiones normativas sobre el significado y el deber de nuestras acciones humanas.

Ignorar estas cuestiones normativas no es una buena solución. Porque, al igual que hay opciones invisibles en los programas digitales, hay normas implícitas que guían nuestras acciones humanas. Ahora bien, nuestro don de la razón nos lleva a creer que la transparencia sobre nuestros objetivos y normas es mejor que la opacidad o no transparencia. Por lo tanto, nos vemos obligados a discutir los valores éticos y a ponerlos en debate.

En el mundo digital, esto sigue ocurriendo muy poco. La iniciativa Cyber Valley, la mayor asociación europea de investigación sobre inteligencia artificial (IA), de la región de Stuttgart-Tübingen, ya ha creado talleres de ética para los investigadores de IA y les exige que discutan explícitamente las consecuencias éticas y sociales de su trabajo. Sin embargo, hasta ahora se trata de una excepción.

Pero incluso en ese caso se plantea la cuestión de qué ética se aplica. ¿Tiene más sentido seguir el imperativo categórico de Kant? ¿O no es más pragmático buscar un enfoque utilitario que priorice la mayor felicidad de la mayoría? ¿O existe la posibilidad de recuperar la ética clásica de las virtudes y las actitudes, por ejemplo, en el marco de una nueva ética del equilibrio?

Lo que se busca, entonces, es el equilibrio adecuado entre las diferentes demandas. No obstante, esto presupone una imagen del ser humano orientada hacia sus mejores posibilidades, no hacia su potencial destructivo; hacia su búsqueda de la buena convivencia, la felicidad y la paz, no hacia los motivos individuales y egoístas.

Este libro presenta una caja de herramientas éticas y analiza los métodos de la toma de decisiones éticas. Se pregunta por la responsabilidad digital y reclama una ética profesional digital también para los programadores y actores del mundo digital. Busca la equidad digital de las empresas y organizaciones, y aboga por la *ética por el diseño*, es decir, por la consideración de las cuestiones éticas ya al principio de la programación.

Por último, tenemos la cuestión de la religión digital. La pandemia del coronavirus ha trasladado muchos servicios de culto y prácticas religiosas al espacio digital. Sin embargo, se trata de nuevas y viejas cuestiones: la inmortalidad digital y la creación digital de seres humanos, pseudopersonas y cuasi-personas. Resulta especialmente apasionante la cuestión de si concedemos dignidad a las máquinas de forma similar a los seres humanos. ¿Existe, pues, una dignidad de las máquinas análoga a la dignidad humana? ¿Qué responsabilidad tienen los humanos con sus criaturas digitales?

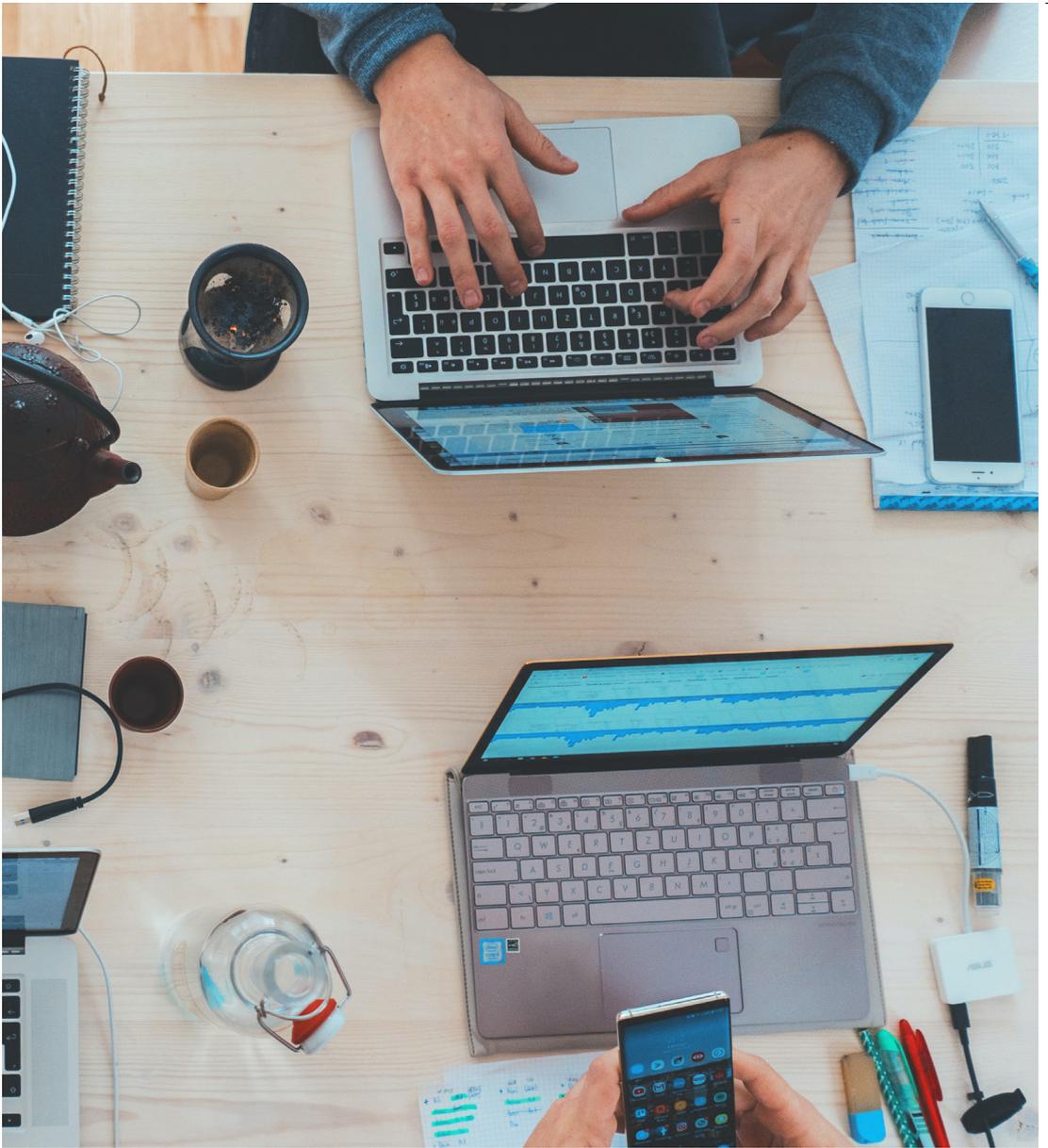
Los humanos solo podemos pensar y actuar como humanos. Si es así, debemos actuar lo mejor posible, aun en el mundo digital. Es tarea de nuestra generación acompañar éticamente el progreso de la transformación digital y aclarar, una y otra vez, de qué se trata al final. Porque la estrella que guía nuestras acciones, también en el mundo digital, debe ser la humanidad. El objetivo es la humanidad digital.

Es un reto permanente buscar soluciones sobre cómo hacer realidad este objetivo en un mundo de grandes tensiones entre diferentes modelos políticos y éticos. Hasta la fecha, no sabemos si los humanos podemos cumplirlo. Nuestra tarea es intentarlo.

Ulrich Hemel
Laichingen y Tübingen,
16 de abril de 2021



1



Ética digital

En el curso de la argumentación en esta serie de libros, hemos abarcado una gran variedad de asuntos sobre un tema que engloba la vida cotidiana, el mundo profesional y el mundo político contemporáneos. Sin embargo, ¿a dónde nos dirigimos? ¿Cuál es nuestro norte? ¿En cuál dirección? ¿Qué queremos lograr? ¿Qué queremos evitar?

La cuestión de una ética digital (cf. P. Grimm, T. Keber, O. Zöllner 2019) en el fondo se dirige a la cuestión fundamental del destino, la finalidad y el objetivo mismo del hombre. La filosofía y la religión se han ocupado de ello durante siglos. A continuación, me gustaría explicar que necesitamos criterios de contenido, no solo formales, para la ética aplicada. A pesar de la apariencia de vaguedad, el concepto de humanidad es adecuado como estándar general.

Julian Nida-Rümelin y Nathalie Weidenfeld desarrollan este estándar en su libro *Humanismo digital* como el núcleo de un enfoque del mundo digital orientado a los valores (J. Nida-Rümelin, N. Weidenfeld 2018). Sarah Spiekermann va en una dirección muy similar, apoyándose en la ética de los valores de Max Scheler (S. Spiekermann 2019).

Max Scheler (1874-1928) comprendió los valores sobre la base de un personalismo ético y su propia forma de análisis fenomenológico (M. Scheler 1913/1916). Desde su perspectiva, los valores se

aplican independientemente de las personas (*cf. M. Scheler 2014*). Este pensamiento permite su discusión, pues los valores no existen independientemente del pensamiento humano, así como de las acciones.

El universo tiene su orden con leyes naturales que solo se interpretan en el contexto de la reflexión humana sobre los valores mirando, por ejemplo, al valor de la supervivencia, la sostenibilidad o la preservación de la creación. Un horizonte de valores como “sostenibilidad” o “preservación de la creación” es, entre otras, la expresión de una cosmovisión religiosa, ya que el concepto de creación presupone un creador.

La ética, y por lo tanto la ética digital, se refleja en este capítulo en un sentido filosófico, no religioso o teológico (*cf. A. Anzenbacher 2003*). Esto corresponde al principio de la “pobreza” o de la economía en hacer suposiciones, axiomas o condiciones previas, es decir, el intento de ser parco con los supuestos anteriores. La ética como disciplina de la filosofía aplicada se dirige a la cuestión de cómo debemos actuar los seres humanos. Por ende, se trata de la evaluación de acciones basada en valores.

Incluso con tal declaración, la ética digital se enfrenta a nuevos desafíos. Si vinculamos los valores y, en consecuencia, la ética con el contenido de la conciencia humana, ¿qué ocurre con la subdisciplina emergente de la ética de las máquinas? Y si la valoración de acciones incluye la atribución de responsabilidad, ¿quién es responsable de qué? ¿El comportamiento de un robot de cuidados, como lo hemos discutido en el capítulo sobre trabajo digital (tomo 2), cuenta como una acción o como un resultado de la máquina? ¿Y cuáles son las consecuencias de tales atribuciones y asignaciones?

No obstante, para la mayoría de los profesionales del mundo digital, de los ingenieros, administradores de empresas y programadores, las

cuestiones éticas, y especialmente las sutilezas de una discusión filosófica especializada, no forman las cuestiones centrales de su propio mundo profesional. Más bien, actúan como “variables de disturbio externo”. Si alguien quiere programar una aplicación que tenga un propósito específico, está buscando una ruta de solución tecnológicamente factible, quizás también a financiadores para la realización económica.

Las consecuencias éticas y sociales se excluyen inicialmente o solo se tienen en cuenta en la medida en que el cliente lo requiera. Este fue el caso, por ejemplo, de la aplicación de advertencia de coronavirus desarrollada por el Gobierno federal alemán, que insistía en considerar las leyes sobre la protección de datos aunque se encontraran en tensión con la idea de la advertencia del virus.

El arte de la argumentación ética como tarea en el mundo digital

Por esta razón, para la ética digital es importante reflexionar, como un punto independiente, sobre el hecho y la práctica de que las cuestiones éticas muchas veces no se discuten, sino se ignoran. Así como es éticamente indiscutible que cualquier omisión es, también, una acción con su responsabilidad correspondiente, esto se aplica a la abstinencia ética de los profesionales en diversos campos digitales: la exclusión de cuestiones éticas y sociales es una posición ética independiente.

Sin embargo, esta posición de abstinencia ética generalmente no es una toma de posición consciente. Es por ello que la práctica digital refleja principalmente lo que parece darse por contado, incluso con los prejuicios en el propio mundo profesional. Si se trata del mundo de los jóvenes blancos de entre 25 y 45 años, su cosmovisión tiene un efecto desproporcionado en los productos digitales.

Sarah Spiekermann lo muestra con el ejemplo ficticio de una aplicación para un servicio de mensajería que entrega comida de

restaurantes. En un segundo paso, genera preguntas sobre las consecuencias sociales y éticas de dicha aplicación. ¿Hay valores implícitos que están programados en la aplicación, como la eficiencia de la entrega? ¿Refleja adecuadamente los valores y necesidades humanos, como la necesidad de los conductores de mensajería de tomar descansos? ¿Qué consecuencias tendrían las decisiones de valor alternativo para la programación de la aplicación? La discusión de tales preguntas conduce a cambios en los objetivos y a nuevos desafíos para la funcionalidad técnica de la aplicación del servicio de mensajería (S. Spiekermann 2019, 24-35).

Las discusiones sobre la nueva “ética de las máquinas” emergente (cf. C. Misselhorn 2018), pero también alrededor de las fantasías sobre cibernets, superinteligencia o “singularidades” (cf. N. Bostrom 2016), muestran que una profunda revolución técnica no puede ser dominada sin cuestiones sociales y éticas.

Para los ingenieros e informáticos en los campos de aplicación digital, una reflexión ética profesional es un desafío. Para los problemas técnicos, a menudo hay soluciones relativamente claras o, al menos, alternativas de solución que dependen de la ruta. En las discusiones éticas, los resultados rara vez son claros, los prerrequisitos a menudo no son ciertos, y los objetivos no están bien definidos. Además, existe el uso del lenguaje metafórico y el salto de un nivel de problema a otro completamente diferente o el salto de una perspectiva a su opuesto.

Con tales observaciones, es fácil pasar por alto el hecho de que la búsqueda de un marco de referencia adecuado, de criterios sensibles y de valores y objetivos con los que se pueda conectar, en sí mismo, constituye una parte considerable de la ética profesional.

Por otro lado, la lucha por el comportamiento correcto en la vida humana choca repetidamente con sus límites. Incluso el mandamiento

“No matarás” (mejor traducido como “No asesinarás”) llega a un límite cuando se trata de defensa propia, ayuda de emergencia o asesinato de tiranos. Hay situaciones en las que el mandamiento práctico “No matarás” no se aplica en la vida cotidiana. En terminología ética, se habla de epiqueyas: la violación justificada de las reglas en una situación ética especial.

Un ejemplo clásico de una epiqueya justificada es que alguien se pase un semáforo en rojo para llevar a su mujer embarazada al hospital a tiempo cuando el saco amniótico ya se ha reventado y el parto es imperativo.

Estos estudios de casos pueden ser comprensibles, pero también conducen a controversias éticas. Esta es una de las razones por las cuales la reflexión ética sigue siendo un desafío enorme, especialmente para las personas en profesiones técnicas y económicas. Hay que observar, entre otras cosas, la confusión que nace del encuentro –a veces incomprensible– entre mundos filosóficos distintos en materia de ética y moral. Además, las controversias éticas suelen ser muy emocionales, lo que va en contra de las afirmaciones de racionalidad de la tecnología y los negocios.

Después de todo, en casi todas las cuestiones y posiciones éticas hay pros y contras, partidarios y opositores, independientemente de que se trate de la pena de muerte, el aborto, el suicidio asistido, la ingeniería genética, la protección de datos o la inteligencia artificial (IA). Muchos de los involucrados reaccionan basándose en su ética privada, es decir, en las normas características que han notado en su crianza y han adoptado para sí mismos.

En muchos casos, no obstante, la actitud de quedarse en el área de la ética privada va acompañada de limitaciones en el arte de la argumentación ética. Esto puede entenderse como la capacidad de revelar los prerequisites para los juicios éticos y la capacidad y voluntad

de cambiar de perspectiva al sopesar los estándares éticos. Por lo tanto, es una competencia que se puede enseñar y aprender porque contribuye al propio juicio ético y a la capacidad de actuar.

Esta habilidad lingüística ética es una competencia central del futuro si queremos vivir juntos pacíficamente en sociedades complejas y organizadas de modo digital. El arte de la argumentación ética requiere familiaridad con los métodos esenciales de la toma de decisiones éticas. Se basa en el conocimiento, la práctica y el uso de un vocabulario ético básico, e incluye actitudes como la voluntad de entablar un diálogo, la apertura al aprendizaje y la capacidad de tomar una posición.

Cada persona tiene su propia perspectiva valiosa debido a su contexto sociocultural y profesional. De otro lado, debe tomar en cuenta que la propia perspectiva es propensa a los errores, que su propia visión puede ser unilateral y que hay perspectivas importantes que, a pesar de todos los esfuerzos, no entran en su campo de visión. Como resultado, todos tienen el derecho expreso a poseer su propia perspectiva y punto de vista.

Esto se aplica, en todo caso, en la medida en que también perciban otras perspectivas y puntos de vista en el diálogo, los tomen en serio y, en caso de duda, los acepten. Las habilidades lingüísticas éticas en la sociedad civil democrática se vuelven, de esta forma, parte de una capacidad de discurso más general, que es particularmente importante en los campos de la ciencia, la tecnología y los negocios. Por el contrario, se postula que la absolutización de un punto de vista individual no puede ser parte del lenguaje ético y la competencia de acción en el sentido aquí sugerido, ni siquiera con miras a las aplicaciones en el mundo digital.

Esta aclaración es fundamental con respecto a los valores religiosos. Al fin y al cabo –como ya se mencionó–, la mayoría de la gente en la Tierra profesa una de las grandes religiones. Tres mil millones de ellos son musulmanes o cristianos. El manejo responsable y reflexivo de las



El arte de la argumentación ética requiere familiaridad con los métodos esenciales de la toma de decisiones éticas. Se basa en el conocimiento, la práctica y el uso de un vocabulario ético básico.

habilidades lingüísticas éticas conoce la amplia gama de posiciones religiosas y no religiosas, pero se distingue de la instrumentalización radical y fundamentalista de ciertas ideas cuando estas no permiten aplicar una opinión diferente.

Parte de la convivencia democrática es la actitud de otorgarse mutuamente el “derecho al error” sin obligar a los demás a aceptar el punto de vista por la fuerza (cf. U. Hemel, 2017, 65-78). Se busca, por consiguiente, una filosofía o una religión que sea capaz de aceptar las reglas de la democracia, pero también una vida pública que sea capaz de aceptar la realidad religiosa. Efectivamente, la búsqueda de la verdad en el sentido de una pretensión de método y de convivencia (cf. U. Hemel 2020) debe ir de la mano con una actitud de humildad que tome nota de la propia posibilidad de error y la combine con la lucha por el mejor conocimiento posible (cf. U. Hemel 2017, 65-78).

El lenguaje ético y las habilidades de diálogo, por ende, forman juntos la actitud de una apertura fundamental al mundo, que es necesaria, además, para la evaluación ética de los fenómenos en el mundo digital. Una tal visión abierta, a veces llamada cosmopolita, es la base de una ética digital humana.

La humanidad en el contexto de la dignidad humana y los derechos humanos

Dado que un punto de referencia común para la ética es un gran desafío, ocasionalmente se adopta la opinión de que las cuestiones éticas no pueden ser generalizadas o “universalizables”. Los juicios éticos se asignan al ámbito de la opinión, así que, supuestamente, tienen un carácter más subjetivo. Tal actitud puede llevar a un rechazo prematuro de las preocupaciones éticas en las aplicaciones digitales.

Un argumento tan escéptico no reconocería el núcleo común de la ética normativa en las culturas, religiones y teorías éticas. Las personas están conectadas entre sí. Sus historias a menudo se entrelazan a través de la composición genética o de historias familiares, pero también mediante tradiciones culturales y religiosas.

En el nivel más alto de abstracción, el objetivo de las demandas éticas es siempre el equilibrio entre el bienestar individual y el bien común. Ese equilibrio es difícil de encontrar. Hay conflictos de intereses, luchas y guerras, pues aun la cuestión del punto de referencia para el “bien común” encierra algunos escollos: ¿Es tu propia familia? ¿Tu propia comunidad? ¿Tu propio país? ¿Europa? ¿Es realmente la totalidad de las personas que viven aquí y ahora en la Tierra, en el sentido de una “sociedad civil global”? ¿O quizás, más allá de eso, las personas de las generaciones venideras, ante las que ya deberíamos sentirnos responsables hoy, en el sentido de la tríada de “medio ambiente natural, medio ambiente social y posteridad”?

¿A qué “nosotros” nos referimos cuando hablamos del bien común? En otras palabras, ¿cuál es el “nosotros” que vale? Porque una cosa es exigir normas éticas mínimas y otra, muy distinta, respetarlas en la práctica diaria, también hacia las minorías y, en general, “hacia los extraños”.

La forma concreta de lo que debe considerarse el objetivo ético primordial de una acción depende de a quién consideramos “persona humana”. Depende, igualmente, de la evaluación de qué grupo de referencia, que es importante para nosotros, debe tener prioridad.

Demos un paso atrás en la historia. Si los bárbaros no son humanos, no necesitan ser tratados con humanidad. Cuando el dominico y más tarde obispo Bartolomé de las Casas (1484-1566) experimentó las atrocidades de los conquistadores españoles contra los indígenas latinoamericanos, se convirtió en un feroz crítico de la conversión forzada y el trabajo forzoso. En el decreto de Burgos (1512) del rey español Fernando II, se prohibió el trabajo forzoso de niños menores de 14 años. Las Casas fue denunciado como un propagandista antiespañol, pero le dio alma a la población indígena de América Latina y, en consecuencia, dignidad humana. Como resultado, la legitimidad del trato increíblemente cruel de estas personas se derrumbó, así como lo manifiestan ciertos edictos reales posteriores (*cf. Th. Eggenberger, U. Engel 1992*).

Desafortunadamente, varias formas de racismo y de actitudes hostiles hacia ciertos grupos humanos florecieron en muchos países hasta en el siglo XXI. Las personas con raíces africanas en Estados Unidos y Alemania, por ejemplo, pueden decir un par de cosas al respecto. Y una y otra vez existe la tentación de ver la cultura de la mayoría de la población como la medida de todo, como en la Alemania nazi, pero también en Japón, durante la Segunda Guerra Mundial, e incluso hoy en China, Irán, Hungría, Estados Unidos y otros países. El poder de los más fuertes parece anular los derechos de los más débiles y de las minorías.

Es tanto más importante que ya exista un criterio para la formulación de estándares mínimos de humanidad, éticos y globalmente efectivos, ya que los seres humanos estamos unidos por el concepto de la dignidad

humana común, por muy diferente que se pueda interpretar (cf. *H. Bielefeldt 1998*).

El concepto de dignidad humana encontró su eco concreto en la Declaración General de Derechos Humanos de la ONU del 10 de diciembre de 1948 y en numerosos textos legales y constituciones, incluida la Ley Fundamental alemana (cf. *P. Tiedemann 2006*).

Aunque la realización de los derechos humanos en contextos religiosos no siempre está garantizada, también en el cristianismo y el islam –las dos religiones más extendidas en la Tierra–, por ejemplo, hay un discurso diferenciado de derechos humanos (cf. *H. G. Ziebertz 2010, M. Pirner, J. Lähnemann, H. Bielefeldt 2016*).

Además, la idea de la ética global, que se remonta al teólogo suizo Hans Küng (1928-2021), busca la comprensión de un conjunto común de valores para todas las personas (cf. *H. Küng 1990, U. Hemel 2019a*). Esta idea fue apoyada por una declaración sobre los valores comunes de todas las religiones formulada por el Parlamento de las Religiones del Mundo (cf. *K. J. Kuschel 2019, 39-54*).

El concepto de dignidad humana, por lo tanto, sirve de norte para la interpretación de lo que se puede expresar con el concepto de humanidad en un sentido normativo. Designa la punta de una pirámide de valores que se despliega de diversos modos en culturas, religiones y formas de ética normativa. El término “pirámide de valores” representa el hecho de que las personas ordenan sus valores jerárquicamente de acuerdo con su importancia y significado. Esto también se aplica si estos valores siempre se interpretan según la urgencia y los requisitos del contexto.

La interpretación del valor máximo y del norte de la “humanidad” en el contexto de la dignidad humana y los derechos humanos va de la

mano con la agotadora variedad de interpretaciones que, incluso, pueden resultar contradictorias. Sin embargo, el concepto de dignidad humana es adecuado como plataforma común para todos los seres humanos, aun en la era digital, pues proporciona un punto de referencia y un marco de orientación que se aplica a todas las situaciones y contextos.

Hoy no solo se tratará de comportamiento humano, sino también de la demarcación de los humanos tanto de los animales como de las máquinas. No obstante, no profundizaremos más en detalle la discusión sobre la demarcación entre humanos y animales, dignidad humana y dignidad animal (cf. R. Hagencord 2011, F. De Waal 2015, U. Lüke, G. Souvignier 2020).

La pregunta adicional de si podemos atribuirles humanidad a las máquinas solo se planteará en el Capítulo 8. Para la cuestión de la ética digital y la ética de las máquinas, empero, es de crucial importancia que tengamos un estándar común con los términos dignidad humana y humanidad –interpretados en el contexto de los derechos humanos–, que sea adecuado como faro y valor de referencia más allá de todas las formas de la ética normativa.

Esta conclusión es válida incluso si se extiende más allá de la especie humana a los animales y las máquinas, precisamente en el sentido de la dignidad animal y la dignidad de las máquinas.

La humanidad, los robots y las leyes de Asimov

El estándar de “humanidad” tiene la ventaja y la desventaja de muchos términos generales que, en vista de interpretaciones divergentes, pueden convertirse en fórmulas vacías. Aquí también hablamos de términos “ómnibus” porque todos pueden llevar su propio bagaje intelectual con ellos en el “ómnibus”.

Sin embargo, una ética digital no ganaría nada con ello. Por consiguiente, es esencial examinar más de cerca el contenido de la norma mencionada con miras a las diferentes tradiciones espirituales de la ética, pero con miras, además, a la relación entre el hombre y la máquina.

Típicamente hoy en día, en el campo de la ética normativa, se hace una distinción entre el utilitarismo, la ética deontológica de Kant y la ética de la virtud (por ejemplo, C. Misselhorn 2018, 53-69). Los diferentes énfasis de estas escuelas de pensamiento solo se considerarán en la próxima sección con respecto a su impacto en la ética digital.

En las siguientes consideraciones, el objetivo es utilizar el criterio de “humanidad” a fin de proporcionar algún ímpetu para la implementación práctica de valores y normas en aplicaciones digitales, pues diferentes conceptos éticos van directamente a la programación de software y hardware, aplicaciones y otros programas. Por ende, es urgente y legítimo que los profesionales digitales exijan una mayor claridad en la orientación ética y, como se ha dicho, un mínimo de habilidades lingüísticas éticas.

En sentido literal, lo fundamental es la visión subyacente de la persona. Si, como expresamente lo formuló Arnold Gehlen (1940/1983), los humanos son vistos como seres deficientes y medidos en función de sus déficits, entonces, en el mundo digital, los humanos son enfocados con sus fallas y límites. La capacidad de atención o vigilancia humana es limitada, la resistencia de las personas está condicionada, y su interés no es constante.

Con tales consideraciones, terminamos en el campo de la búsqueda de ayudas para expandir nuestros sentidos y nuestras capacidades cognitivas, es decir, en el campo de la búsqueda de prótesis. Más allá de las gafas y los audífonos, los microscopios y las ecosondas, el mundo digital trata de aumentar las capacidades cognitivas.

Las consecuencias de esto son, entre otras cosas, las utopías de mejora (S. Dickel 2011), o sea, las utopías del incremento del rendimiento de frente a la implacable afirmación de la autooptimización. Si se limita el valor de la persona humana a su rendimiento cognitivo o general, en el caso extremo distópico llegamos a la idea de que los cibernéticos y las superinteligencias son, por así decirlo, las “mejores personas”. Por ejemplo, el infinito y la vida eterna se lograrían descargando mi cerebro en un soporte de datos y un procesador adecuados.

Esta es la dirección en la que van algunos juegos de pensamiento transhumanistas, que se discutirán con más detalle en el próximo capítulo. No obstante, se trata de la cuestión de la ética digital con miras a la perspectiva elegida sobre imágenes de la persona humana y sobre normas éticas. Es parte de la peculiaridad de los seres humanos que, en oposición a la idea de una existencia concebida cognitivamente, existen facetas y experiencias muy distintas de la vida biológica, espiritual, psicológica y social que fluyen entre sí en personas humanas concretas.

En la “crítica de la razón digital” presentada aquí, se asume que la “humanidad” también puede representar una brújula de primer rango para la ética digital. El concepto de humanidad, por lo tanto, se entiende en un sentido normativo, lo que se basa en las mejores tradiciones humanas de tratar con otras personas y grupos de personas.

Por tal motivo, en primer plano no se halla la posibilidad empírica de la maldad, el desprecio por los seres humanos y el crimen sino, más bien, la pretensión humana de estar a la altura de sus mejores capacidades, es decir, la “humanidad” más allá de la afiliación y distinción étnica, religiosa, sexual, social o cultural.

La humanidad se convierte así en el norte y la vara de medir para interpretar las demandas éticas. Los contextos de esta interpretación serán diferentes, así como las conclusiones concretas. Sin embargo,

Cuando se trata de funciones cognitivas o sensoriales, de acuerdo con el estado actual de la técnica, las máquinas pueden superar a las personas en términos de rendimiento, resistencia y velocidad.



es útil tener un “norte” y buscar criterios comunes más allá de toda controversia. Por un lado, esto crea transparencia sobre posibles diferencias de valoración. Por otro lado, una búsqueda seria y conjunta revela la lucha fundamental por el mejor camino. Dicha actitud de búsqueda y lucha es, en sí misma, una expresión de humanidad, de voluntad de moldear, y no una de destruir.

El estándar normativo de humanidad va más allá de las leyes de Asimov, formuladas por Isaac Asimov (1920-1992), porque sus leyes solo se refieren a personas empíricamente existentes y, en el mejor de los casos, a la “humanidad” en sus contradicciones vivas. No exigen el valor normativo que se entiende por “humanidad”.

Mientras tanto, la mera referencia a la especie humana y su superioridad declarada ha sido criticada como “especismo” (P. Singer 1982). El polémico veredicto del “especismo” podría llevarse más lejos en el mundo digital, extendiendo el discurso sobre la dignidad humana y los derechos humanos a los robots o cíborgs, en el sentido de “dignidad de las máquinas” y “derechos de las máquinas”. Pero el hecho de que los requisitos normativos para evitar el “especismo” se originen únicamente en los seres humanos y estén orientados solo a la acción humana puede verse como una contradicción a tal tesis del “especismo”.

No obstante, las leyes de Asimov –originalmente formuladas en un relato corto en 1942– se relacionan con las acciones de los robots, no directamente con las de los programadores humanos o quienes tomen las decisiones. Al describir la relación hombre-máquina, las leyes de Asimov recuerdan más la dialéctica de “amo y sirviente” y las estructuras sociales de los siglos XIX y XX. Cuando se dice “no hagas daño a nadie” u “obedece las órdenes de los humanos”, los robots tienen que respetar a los humanos como una especie superior porque se supone que es parte de su programación.

En aquel entonces había incluso preocupación por la superioridad incontrolada de los robots. Esto se puede contener si los seres artificiales se adhieren a las leyes que dan prioridad a las personas. Así, la especie humana tendría que ser respetada dentro del marco de las leyes de Asimov, pues tiene mayor poder a través del acceso a la programación de los robots. El argumento del poder puro como base de la superioridad afirmada es, éticamente, de alcance muy limitado.

La superioridad implícitamente afirmada de la especie humana ya está sobre pies de barro en el periodo temprano de una especie de “ética de las máquinas”. Cuando se trata de funciones cognitivas o sensoriales, de acuerdo con el estado actual de la técnica, las máquinas pueden superar a las personas en términos de rendimiento, resistencia y velocidad. En lo que respecta al poder, son muy concebibles las constelaciones que doten a las máquinas de un poder superior al de los humanos.

Cuando se trata de una superioridad que se ha vuelto históricamente importante, todos saben que la historia es impredecible y puede cambiar su curso de forma dramática. La crisis del coronavirus del 2020 –que nadie esperaba– lo muestra claramente. Si, en realidad, los humanos vamos a tener una función de control superior, esto debería alimentarse de la fuerza moral de los humanos y no referirse a las actuaciones físicas o cognitivas individuales.

La fuerza moral no significa que las personas no puedan cometer errores. Por el contrario, la fuerza moral es el resultado de orientar las acciones humanas hacia el criterio de “humanidad”, aunque este reclamo no se cumpla en la acción concreta. Por consiguiente, las personas pueden lograr o no lograr este objetivo, pero es y sigue siendo el punto de referencia. Además, es parte del estándar normativo de humanidad que las personas aprendan de los errores y puedan corregirlos. Se trata de una especie de “capacidad de autocontrol” de segundo orden, por lo que hasta cierto punto equivale a una ecuación diferencial ética.

Considerándolas en detalle, las leyes de Asimov son las siguientes (I. Asimov 1982, 67):

1. Un robot no hará daño a un ser humano ni, por inacción, permitirá que un ser humano sufra daño.
2. Un robot debe cumplir las órdenes dadas por los seres humanos, a excepción de aquellas que entren en conflicto con la primera ley.
3. Un robot debe proteger su propia existencia en la medida en que esta protección no entre en conflicto con la primera o con la segunda ley.

La primera ley a veces se complementa con el adjetivo “a sabiendas”, ya que solo se pueden abordar las infracciones “a sabiendas”. Si reflexionamos sobre dichas leyes a la luz de las consideraciones anteriores respecto a la humanidad, queda claro que las reglas del juego formuladas de ese modo, a pesar de la regla 2, también protegen a los criminales de las máquinas, siendo los delincuentes personas humanas.

Sin embargo, si consideramos la esfera humana, serían excluidos ejemplos extremos, como la autodefensa o el tiranicidio, hasta en los casos en que el objetivo sea evitar un desastre aún mayor. Además, es importante discutir cómo se debe interpretar la regla mencionada para

los dilemas éticos. Estos tipos de conflicto se discuten intensamente en la ética de las máquinas, así como, por ejemplo, en casos de conducción autónoma y de drones militares.

Sobre la base de esas consideraciones, las leyes de Asimov se complementaron con una ley “cero” de la siguiente manera: “Un robot no debe dañar a la humanidad o, por pasividad, permitir que la humanidad sea dañada”. Empero, si tal ley se aplicara realmente, un ser humano individual muere cuando se evita el daño a la humanidad. Sigue siendo problemático el hecho de que la cantidad y el tipo de daño no se definen con precisión aquí. Éticamente, la cuestión de la responsabilidad y cómo lidiar con el poder surge de manera muy aguda.

La persona que actúa en defensa propia rinde cuentas, y posiblemente sea absuelta. Sus acciones son objeto de una mayor atención social, ya que traspasan los límites éticos normales, por ejemplo, en la forma de un proceso judicial. En la conocida película *Yo, Robot* (director: Alex Proya, EE. UU. 2004), los robots controlados por el sistema de software VIKI (Inteligencia Artificial Virtual) toman el control de las personas para protegerlas (cf. *J. Nida-Rümelin, N. Weidenfeld 2018, 64-70*). Pero, dependiendo de la situación, esto significa que las personas individuales tienen que pagar con su vida para proteger a la humanidad. Esto es parte de la programación, y no se considera un “manejo excepcional” que se comprueba de modo específico.

De igual manera, las aplicaciones militares –entre ellas, los drones– deben considerarse intrínsecamente como violaciones de, al menos, tres de las leyes originales de Asimov.

En la práctica, no obstante, el uso de drones militares se justifica y se hace cumplir recurriendo a un “bien superior”, como la protección contra la tiranía y el genocidio. Quien aquí tiene el poder puede, empero, hacer valer su punto de vista y sus intereses frente a argumentos controvertidos.

Pese a ello, la ética digital rara vez es tan dramática. Es inevitable que las decisiones deban tomarse sobre la base de juicios de valor implícitos o explícitos. Cómo surgen y cómo pueden justificarse es el tema de la siguiente sección.

Ética utilitaria, deontológica y de la virtud: la humanidad como criterio

Las personas viven su mundo como una secuencia de situaciones que tienen el efecto de estimular acciones humanas. Esto incluye procesos internos, como el hambre o la sed, pero también situaciones sociales, como en el tráfico o en el trabajo. Cada situación tiene parámetros de decisión explícitos e implícitos que contienen juicios de valor ético, aun si esto no es reconocible de inmediato.

Este es un gran desafío para el mundo digital cuando se trata, por ejemplo, de aplicaciones no discriminatorias. El término “discriminación”, en el sentido de trato injusto a grupos de personas a los que se les asignan determinadas características, oscila entre la discriminación deliberada y no intencionada. Si la discriminación ocurre involuntariamente por la falta de conciencia de una determinada minoría, esto da como resultado un vago sentimiento de inseguridad para el programador individual: puede pasar por alto algo que ni siquiera nota.

Además, el concepto de discriminación siempre depende de interpretaciones en el ámbito social y político. Lo que ayer se consideraba normal puede verse hoy como discriminación. Hasta los años 60 del siglo pasado, por ejemplo, la forma correcta de dirigirse a las mujeres solteras en alemán era “*Fräulein*” (señorita). Hoy en día, el término casi se ha extinguido y se percibiría como extremadamente discriminatorio.

Por las razones mencionadas, los programadores toman decisiones éticas incluso cuando no quieren o ni siquiera se dan cuenta. Si es así,

tiene sentido sopesar diferentes razones éticas entre sí y, por ende, es lógico que las historias comunes de la ética normativa puedan volverse relevantes en el contexto de la programación digital. Por ello, las siguientes explicaciones pretenden arrojar luz sobre las direcciones principales, es decir, la ética utilitaria, deontológica y de la virtud. La ética clásica de la virtud debe desarrollarse más en la dirección de una ética de los valores y una ética del equilibrio, porque esto abre un acceso éticamente sólido, pero también pragmático, a las decisiones éticas en el mundo digital.

En la región anglosajona, con su gran tradición de búsqueda individual de la felicidad, la ética utilitaria está particularmente extendida. Esto sigue el estándar de la mayor felicidad del mayor número formulado por Jeremy Bentham (1748-1832) (*cf. J. Burns 2005, 46-61*). La ventaja de la idea utilitaria radica en que es relativamente fácil de poner en práctica, ya que es muy probable asignar un número en una escala ordinal a varios efectos. Si uno ignora los problemas filosóficos detrás de cada teoría de medición, se llega a valores numéricos para la conveniencia de decisiones éticas que pueden ser fácilmente representadas de manera digital.

La supuesta buena capacidad de operación de las decisiones éticas en un enfoque utilitario omite, sin embargo, una serie de problemas éticos graves. Cuatro de ellos se nombrarán a continuación.

La exclusión de las decisiones sobre mensurabilidad de beneficios debería ser vista críticamente. Incluso la elección de la escala puede ser un problema metodológico porque los valores numéricos en una escala de 0 a 5, en lugar de una escala de 1 a 3 o de 1 a 10, dan imágenes muy diferentes. Dichos temas se analizan en detalle en la discusión metodológica de las ciencias sociales (*cf. E. Roth, H. Holling 2018*).

El siguiente nivel de dificultad de aplicación radica en el horizonte temporal de beneficios. ¿Existen beneficios a corto, mediano o largo plazo? ¿Cómo se definen exactamente a corto, medio y largo plazo? ¿Y quién decide la definición éticamente correcta de un horizonte temporal apropiado?

En tercer lugar, se trata de un análisis complejo de intereses: ¿Qué beneficios intervienen en qué ponderación? En principio, las cuestiones éticas no pueden dissociarse de las cuestiones de poder. No obstante, debido a que los conflictos de poder y los conflictos de intereses también son parte de la realidad, los programadores se sentirían injustamente sobreexigidos si se esperara que sopesen cuestiones de poder e intereses al mapearlos en la ponderación digitalizada de los momentos de toma de decisiones.

Por último, una cuarta cuestión muy compleja se relaciona con el proceso de medición real de los beneficios. ¿Se trata del beneficio de una persona, una organización o una comunidad entera? ¿Existen procedimientos reconocidos para “medir” el beneficio correspondiente? ¿Esta medida también se aplica a todos los horizontes temporales, es decir, a corto, mediano y largo plazo? ¿O solo se muestran “opiniones” sobre “posibles beneficios” al final?

En otro capítulo examinamos el ejemplo de la grabación digital del comportamiento de un anciano diabético residente en un hogar que come chocolate. Entonces, ¿cómo se deben evaluar los beneficios o daños del chocolate? ¿Cómo se evalúa el disfrute que siente el mencionado anciano? ¿Cómo se evalúa el impacto económico del posible deterioro de su salud?

Esta discusión no debe llevarse a cabo aquí en detalle porque ya está claro que la evaluación de beneficios da lugar, inevitablemente, a una determinada posición de poder. Quien tiene derecho a hacer

cumplir su propia evaluación de beneficios ejerce el poder. En lenguaje sociológico se habla aquí del privilegio de la definición.

Eso se vuelve problemático cuando durante la programación de una aplicación digital se supone, de manera implícita o explícita, un cierto beneficio, con consecuencias en las situaciones de uso. Tales consecuencias se pueden considerar como manifestación de un poder decisonal indebido y, por lo tanto, como actos de “promoción o disminución de la humanidad”. Tal reducción puede, por ejemplo, residir en la sensación de control externo, en lugar de una promoción de la libre autodeterminación.

El objetivo de las aplicaciones digitales debería ser acrecentar la humanidad y la libre autodeterminación. Debido a que las consideraciones utilitarias no cumplen inequívocamente con este estándar, en última instancia, no son un estándar adecuado para la ética digital.

En la Europa continental, a diferencia del utilitarismo, la ética deontológica de Kant (1724-1804), es decir, orientada hacia un “debe”, tiene numerosos seguidores. “Obra de tal modo que trates a la humanidad, tanto en tu persona como en la de cualquier otro, siempre como un fin y nunca solamente como un medio” (I. Kant 1968, 429).

El enunciado básico en el trasfondo de esta solicitud es, más allá de la prohibición de la instrumentalización por parte de las personas, la exigencia de una ética del deber que fije el estándar del comportamiento individual a su generalización. “Obra sólo según una máxima tal que puedas querer al mismo tiempo que se torne en ley universal”. Esa es una de las paráfrasis comunes del imperativo categórico de Immanuel Kant.

La ventaja de la ética deontológica es, por un lado, la clara jerarquización del comportamiento a través de la conexión con una ley general. El robo, por ejemplo, es una práctica claramente inadecuada

para la legislación de alto nivel. Una segunda ventaja de la ética kantiana es la primacía de la persona que no debe convertirse en un fin, sino que siempre tiene un valor intrínseco.

El a menudo descrito “dilema del tranvía” de la conducción autónoma tiene un efecto muy práctico, en el que la ética de Kant trabaja en contra de una lógica instrumental. El “dilema del tranvía” se refiere a situaciones en las que no se puede evitar un accidente fatal (*cf. C. Misselhorn 2018, 190*). En tal situación, una vida no puede compensarse con otra.

Los seres humanos no tienen derecho a esa compensación, al igual que las máquinas o los programadores de máquinas. Específicamente, no se puede programar información adicional, como la edad de la víctima potencial del accidente, en la ruta de toma de decisiones de un vehículo autónomo, como si fuera éticamente mejor entregar a las personas mayores a la muerte en un accidente, en lugar de hacerlo con las más jóvenes.

El primer punto a criticar por las relaciones digitales en el nivel práctico es la muy difícil operacionalización del imperativo categórico. El programador medio no es un especialista en ética. El trabajo conceptual filosófico necesario para una aplicación selectiva de la ética deontológica no suele ser, en absoluto, razonable para él.

Además, los expertos en ética profesional no siempre están de acuerdo entre ellos. Otra objeción retoma la cuestión de la viabilidad y la lleva más allá: ¿De qué tipo de leyes debería tratarse y cómo deberían interpretarse? Al final de cuentas, existen numerosas leyes que sirven a grupos de interés muy específicos, pero que se aprueban y se hacen cumplir. Incluso el robo se puede verter en forma de leyes, por ejemplo, en el caso de la expropiación de negocios judíos durante el Tercer Reich.



La ética deontológica es, en última instancia, solo parcialmente adecuada como punto de referencia para evaluar la práctica digital.

La consideración de que tal ley es inherentemente ineficaz o al menos inadmisibles bajo la ley natural, o basada en principios deontológicos, juzga de manera inadecuada la realidad social de la aplicación estatal de las leyes, aun si estas son éticamente cuestionables o, de hecho, deben ser desestimadas.

Ahora se podría argumentar que hablamos de leyes en un mundo ideal. Desafortunadamente, en este contexto también cabe hablar de falta de aplicabilidad en el mundo real. Después de todo, las personas imaginan el mundo ideal de forma muy distinta. Una ética deontológica no puede ir más allá del alcance de una interpretación individual del mundo ideal.

El estándar de la ley general no se aplica a todos los contextos y comportamientos sensibles y éticamente aceptables. Retomemos el ejemplo del residente comiendo un trozo de chocolate. Sabe que le queda poco tiempo de vida; unos días más o unos días menos no le importan. Ninguna ley prohíbe comer chocolate, incluso cuando existan riesgos para la salud.

¿Cuál sería la “ley general”? Es posible generar formulaciones y perspectivas diferentes, pero lamentablemente contradictorias: “Todo el mundo come lo que le gusta”, o “nadie daña deliberadamente su

salud”, o “la autodeterminación es el valor más alto en comparación con las restricciones de salud”, o “el modelo a seguir para los demás también se aplica a los residentes mayores que tienen apetito por chocolate”.

En cualquier caso, queda claro que no se puede esperar que un programador desarrolle su propia posición competente, aunque siempre controvertida, sobre tales cuestiones. Por esta razón, la ética deontológica es, en última instancia, solo parcialmente adecuada como punto de referencia para evaluar la práctica digital.

Un tercer enfoque ético se resume, a menudo, bajo el término general de virtud o ética de actitudes.

Mientras que la ética utilitaria y deontológica con Bentham y Kant se remonta a dos autores del siglo XIX en la Europa cada vez más secular de la Revolución Francesa, la ética clásica de la virtud tiene una multitud de fuentes, algunas de las cuales son muy heterogéneas, en términos de tiempo y tema. Estos van desde la antigua Grecia hasta la Edad Media cristiana. También conviene mencionar los enfoques fenomenológicos de la ética de los valores de Max Scheler (*cf. M. Scheler 2014*). El arco de la ética de la virtud abarca más de 2000 años, y siempre incluye nuevas interpretaciones religiosas y seculares.

A menudo, las raíces religiosas de la ética se pasan por alto en las discusiones sobre filosofía académica. No obstante, estas son efectivas en la vida práctica de millones de personas, y su impacto debería ser considerado igualmente por las personas sin convicciones religiosas (*cf. T. Crane 2019*). Ejemplos de convicciones religiosas generalmente efectivas en el campo ético son la “regla de oro” de la reciprocidad y los Diez Mandamientos bíblicos.

Incluso la ética de las religiones del mundo no cristianas, a pesar de toda su diversidad y diferenciación, no es fundamentalmente una ética

utilitaria o deontológica, sino una ética de valores, actitudes y virtudes. El concepto algo anticuado de virtud es reemplazado, cada vez más, por el concepto de valores que guían la acción y el concepto de orientación al valor (cf. H. Joas 1997). Esto se aplica también al área económica (cf. U. Hemel 2007). Cualquier interesado en una interpretación moderna de “virtud” y “virtudes” puede encontrar esto, por ejemplo, en Otto Bollnow (1903-1991), Josef Pieper (1904-1997) o Alasdair MacIntyre (nacido en 1929).

En el contexto europeo, la ética de la virtud se remonta habitualmente a Aristóteles (385-323 a.C.) o Tomás de Aquino (1225-1274).

El término “ética de la virtud” se refiere, entonces, al carácter de una persona que actúa, es decir, a los contextos éticos individuales. La restricción a la persona individual no es inevitable, sin embargo, porque la orientación en valores y el comportamiento habitual no son solo un asunto de individuos, sino de organizaciones, instituciones y empresas.

Para Aristóteles, el concepto de eudaimonía o felicidad está en primer plano. Si se le interpreta, no en términos de ética individual, sino en términos de ética social, se llega a una orientación hacia el bien común como principio de ética. Una expresión moderna de la ética clásica de la virtud sería la orientación valorativa en el equilibrio entre la búsqueda de la propia felicidad y la búsqueda del bien común.

Esta perspectiva también me parece adecuada para el mundo digital. La he representado desde que tuve ocasión en mi libro *Valor y valores: Ética para directivos* (U. Hemel 2007) para reflexionar sobre mis propias experiencias de gestión de una empresa con miles de empleados con miras a la orientación práctica de los valores. La discusión más reciente sobre propósitos corporativos (*purpose*) o finalidades de crear “sentido” y de contribuir al bien común va en la misma dirección. Puede entenderse como una actualización o como una variante adicional

de la “ética de la virtud” o una ética de orientación de valores (*cf. N. Mourkogiannis 2006, F. Fink, M. Moeller 2018*).

Una ventaja general del concepto de ética de las virtudes es su viabilidad, combinada con transparencia y auditabilidad. Las personas son capaces de aprender y pueden reconocer, de manera autorreflexiva, su propio comportamiento como adecuado o como un error y corregirlo. En todo caso, las acciones repetidas a lo largo del tiempo devienen en actitudes vividas que, en última instancia, dan como resultado orientaciones de valores estables que son o “vicios” o “virtudes”.

El matiz específico de las actitudes valorativas es sociocultural y situacionalmente distinto. De esta forma, dicha ética de equilibrio cumple, en la práctica, con el principio de humanidad.

Veamos un ejemplo práctico. Un padre conduce hasta un lago con su hijo, levemente discapacitado, y los hijos de los vecinos. Los tres niños que juegan resbalan, uno tras otro, en un embarcadero y caen al agua. Para un padre, es una reacción natural, y también éticamente correcta, apresurarse a salvar primero a su hijo y, luego, a los demás. En la ética de Kant, esto sería problemático, ya que actuar por deber no permitiría la preferencia por los propios hijos y los amigos. Para los protagonistas de la ética utilitaria, por otro lado, podría surgir una discusión sobre si no es éticamente necesario salvar a un niño con mayor productividad futura, en lugar del niño discapacitado. Pero incluso una discusión así sería percibida por la mayoría de la gente –y con razón en mi opinión– como una violación del principio de humanidad.

El concepto de ética de la virtud puede ser criticado por su potencial contradicción y su posible falta de validez general debido a que es la interpretación situacionalmente flexible de los valores la que conduce a las contradicciones, pues se cuestiona la consistencia general de la acción. Esto puede verse como un argumento contra la universalidad.

No obstante, pienso que esta crítica, fundamentalmente justificada, no es decisiva porque nosotros, como seres humanos, siempre estamos atados a una perspectiva que influye en nuestra cognición y acción mediante la edad, el idioma, la cultura y la orientación de valores. Es por eso que nunca hay valores vividos sin el color y el tono concretos de un horizonte de interpretación situacional y contextualmente específico.

Además, las personas son capaces de aprender, es decir, pueden cambiar su comportamiento sobre la base de conocimientos más recientes. Y, finalmente, son capaces de usar la razón. Por lo tanto, en las diversas situaciones de la vida distinguen según el caso y saben adoptar criterios de conducta o de acción, así que entienden la razón por la cual es mejor actuar de una manera en ciertas situaciones y de manera distinta en otras.

La capacidad humana de abstraer nos permite reflexionar sobre nuestros propios valores y configuraciones socioculturales y hacerlos transparentes. Por consiguiente, una ética del equilibrio orientada a los valores corresponde, en particular, al principio de transparencia, equidad y posibilidad de corregir errores.

En consecuencia, no se debe reclamar la universalidad donde, por ejemplo, la perspectiva especial de los hombres blancos mayores se considera universalmente válida. Más bien, es justo y transparente reflexionar sobre la propia perspectiva como una oportunidad y un límite y, también, incluir otras perspectivas (entre ellas, las de las mujeres negras más jóvenes) tanto como sea posible.

Esta argumentación es de singular importancia para la práctica digital porque los valores de equidad y transparencia están recibiendo especial atención en nuestro tiempo, incluso a nivel global. En otras palabras: la humanidad se interpreta con vistas a la equidad y la transparencia también en el ámbito digital.

Al fin y al cabo, los valores de una ética del equilibrio se pueden aplicar y entrenar bien. No es una desventaja que los valores experimenten cambios con el tiempo que conduzcan a ajustes en el comportamiento deseable y, por ende, en la programación.

Arquitectura mental y contextos situacionales de acción

La ética digital presenta a los seres humanos retos en los que el recurso a las anteriores teorías normativas de la ética llega a sus límites debido a nuevas interrogantes, a la difusión de responsabilidad, a la gran cantidad de actores y a las consecuencias que no se pueden valorar con claridad.

Por ejemplo, cualquier persona que programe una aplicación no solo debe actuar desde la sensación de su propio entorno de vida, sino también debería poder reflejar los efectos en contextos sociales más amplios. Esto requiere de una práctica sistemática de enseñanza y aprendizaje sobre valores y comportamientos que promuevan y no menoscaben la humanidad. Como se describió, la solución ideal para una ética sensata del equilibrio es su interpretación en el contexto de la dignidad humana y los derechos humanos.

La orientación normativa hacia la promoción de la humanidad es una cosa; su justificación biológica, social y antropológica es otra. La cuestión del origen y la validez de los valores (*cf. H. Joas 1997*) no se desprende de la constitución biológica de la especie humana. Es el examen de la representación digital de valores lo que nos lleva a preguntarnos cómo podemos entender la forma específica en que los animales, las personas y las máquinas procesan la información.

Esta pregunta es de gran importancia para una “crítica de la razón digital”, incluso si evitamos las complejidades de la discusión especializada en psicología cognitiva y antropológica (*cf. M. W. Eysenck, M. T. Keane 2010; E. Bohlken, Ch. Thies 2009*). A fin de cuentas, es

obvio que los humanos y los animales procesan información de su entorno (cf. *H. Wagner 2020, 45-65*). El hecho de que la información se procese está a veces relacionado con la teoría de que el cerebro humano debe entenderse de la misma forma que una computadora.

La analogía computadora-cerebro en el cognitivismo (cf. *J. R. Andersson 1983, R. Gaschler, P. A. Frensch 2009, 156-163*) no es de gran ayuda para la ética digital, ya que termina en una especie de razonamiento circular: las personas han creado computadoras con sus cerebros y ahora interpretan el funcionamiento de sus cerebros según el modo de las computadoras programadas digitalmente.

Sin embargo, calcular no es lo mismo que asignar un significado o entender y, ciertamente, no es el entendimiento en un contexto situacional. Las personas son más complejas que las computadoras, tanto por sus diferentes niveles de percepción como por su horizonte temporal y su capacidad para ser libres y autodeterminadas. La ética digital debe tener en cuenta esta complejidad antropológica.

En este punto, me gustaría discutir más en detalle el término de “arquitectura mental” (cf. *U. Hemel 2019, 335-350*). La arquitectura mental de una persona u organización significa una estructura relativamente estable de hábitos de percepción, clasificación interna, reconocimiento de patrones y generación de un cierto rango de reacciones.

La arquitectura mental de una persona tiene, al mismo tiempo, la característica de formarse de manera histórico-biográfica, genéticamente condicionada y adaptada a la situación. En otras palabras, la arquitectura mental representa la totalidad de conexiones significativas que, consciente o inconscientemente, pueden considerarse típicas del carácter y la personalidad específica de un individuo, además de incluir evaluaciones emocionales y éticas.

Nuestra arquitectura mental, en el sentido de recurso interpretativo relativamente estable en el tiempo, pero que puede activarse según la situación, decide la relevancia y plausibilidad de lo que nos sucede en la vida (*cf. U. Hemel 1988, 290-290*). Asimismo, la calidad de nuestra clasificación o formación de categorías (encuadre) determina la dirección e intensidad de nuestras reacciones ante una situación dada (“*afrentamiento*”; *cf. ibid., 289*).

Partimos de la observación cotidiana de que las personas –como los demás seres vivos– tienen que orientarse en su vida cotidiana y en toda su vida. Están diseñadas para tomar decisiones que pueden entenderse como prioridades en la ejecución directa: hacemos una cosa y no la otra. Comemos un trozo de tarta o prescindimos de él; vamos a nadar o leemos un libro; recibimos formación a largo plazo o la dejamos; nos casamos y criamos hijos, o por varias razones no lo hacemos.

Los fenómenos aquí mencionados suelen ser examinados desde aspectos como la percepción, la atención, la memoria o la motivación. Desde el punto de vista de la arquitectura mental, se trata de la forma subjetiva de una teoría unificada, una teoría individual sobre el mundo, la cual comprende una autoimagen y una cosmovisión, un concepto de nuestra propia identidad y un marco para nuestra comprensión del mundo. La idea de arquitectura mental se refiere a la conexión sistémica y sistemática de nuestra existencia en el mundo y nuestra forma individual de acceso al mundo y procesamiento de este (*cf. U. Hemel 1988, 543-564*).

Las percepciones lingüísticas, emocionales y, en un sentido más estricto, cognitivas fluyen entre sí en los seres humanos, pero no en las máquinas. Las máquinas procesan la entrada y entregan la salida. Como regla general, no hay una visión explícita y completa del mundo y del “yo” detrás de esto. La gente vive, y su vida no siempre se orienta hacia la producción funcional.

A diferencia de las máquinas, los humanos se ven obligados a interpretar el mundo en términos del significado para ellos mismos. Igualmente, tienen la tarea inevitable de interpretarse en el contexto de su mundo. Como resultado, no solo obtienen una imagen de sí mismos y del mundo, sino que también forman su propia identidad en la interacción de la estructura genética, el carácter sociocultural y el autocontrol (cf. *U. Hemel 2017, 157-173*). Las personas están obligadas a interpretar el mundo, pero también tienen su propia competencia para interpretarlo (cf. *U. Hemel 1988, 556*).

El concepto de interpretación del mundo que se utiliza aquí está inicialmente relacionado con los eventos biológicos. La gente tiene que orientarse en su mundo, por ejemplo, enfocándose en la comida, el peligro o la afiliación social.

Debido a la neuroplasticidad del cerebro humano, y especialmente a la capacidad humana de hablar, simbolizar y abstraer, la interpretación muy simple del mundo por parte de un niño pequeño se desarrolla dentro de un marco sociocultural dado. Con el tiempo, conduce a un modelo de mundo socialmente adaptado y configurado individualmente en el marco de la arquitectura mental de una persona. Dicho modelo mundial incluye los resultados de la autoeducación y el aprendizaje de la identidad, de asumir roles e incorporarse, de manera que uno pueda decir, por ejemplo: “Soy un centroeuropeo tolerante, de mente abierta, con una buena educación, y estoy particularmente comprometido con mi propia familia”.

Los procesos de identificación religiosa e ideológica también se reflejan en el modelo mundial de tal persona: “Tengo una educación cristiana, pero no estoy muy apegado a la Iglesia” o “No he sido educado religiosamente, y he dejado abierta la cuestión de Dios y de la religión”. La competencia fundamental para interpretar el mundo se expresa, entonces, en el nivel de la propia comprensión de uno

mismo y del mundo, pero también en supuestos psicológicos sobre la propia persona.

No obstante, dependiendo de la situación, puede haber tensiones y contradicciones entre el propio modelo de mundo y la interpretación del mismo, por ejemplo, si alguien está convencido de que mentir no es bueno, pero se deja llevar a la mentira durante un control policial del tráfico. Por lo tanto, se puede hablar de una brecha entre competencia y rendimiento situacional, al igual que un pianista de clase mundial probablemente no tocará un concierto de clase mundial cuando esté borracho y, por ende, no esté en condiciones de manifestar su máximo potencial.

En cualquier caso, competencia y desempeño van de la mano pues la competencia de interpretación del mundo se expresa situacionalmente en una interpretación concreta de aquel (*cf. U. Hemel 1988, 559*). Se trata de la clasificación situacional de lo que experimentamos y nuestra reacción ante él. Si un hombre se para en la cocina con un cuchillo grande y levanta el brazo, podría ser un intento de asesinato o el típico acto de cocinar. Si nos paramos a su lado, activaremos emocionalmente nuestra “competencia de interpretación del mundo”. Los movimientos enojados y agresivos podrían asustarnos; los movimientos relajados y decididos para cortar verduras no serían motivo para una mayor atención.

Sobre todo en el campo de la terapia conductual, la “teoría del marco de referencia” desarrollada por Steven C. Hayes (*cf. F. Binder et al. 2020, 186*) y la terapia de aceptación y compromiso que se remonta a él y sus colegas se han visto utilizadas con éxito (*cf. S. C. Hayes 2001, S. C. Hayes, K. D. Strosahl, K. G. Wilson 2014*). También se trata de la “clasificación”, es decir, del “enquadre” de lo que nos sucede. Por consiguiente, los cambios en las clasificaciones pueden ayudar a obtener una visión diferente de las cosas y, así, evitar un estrés subjetivo excesivo (*cf. F. Binder et al. 2020, 183-202*).

A primera vista, puede parecer sorprendente cuando aquí se alude a investigaciones que originalmente se discutieron en campos como la educación religiosa, la psicología cognitiva o la psicoterapia. En realidad, la transferencia de procesos de clasificación (encuadre) y la construcción de modelos mundiales a aplicaciones digitales se recomienda porque los sistemas basados en IA –como la conducción autónoma– no pueden funcionar sin un diseño correspondiente de un modelo ambiental con procesos comparables de clasificación, reconocimiento de patrones y afrontamiento de la situación. Por ello, las consideraciones sobre la arquitectura mental son tanto ética como digitalmente relevantes.

Volvamos a nuestro ejemplo del “hombre enojado en la cocina”. La relevancia de los modelos del mundo significativos se debe principalmente a que se evitan las percepciones y reacciones incorrectas. Es muy probable que ciertas percepciones erróneas conduzcan a acciones incorrectas; por ejemplo, si reacciono con “calma” por una percepción inadecuada de la situación, pero el hombre de la cocina está realmente enojado y me ataca de forma agresiva.

De modo que existen ciclos de retroalimentación correctiva en la relación entre el mundo y el hombre, así como en la relación entre el mundo y el sistema, que tienen un efecto en nuestro sistema de conocimiento y acción. Sin un mínimo de adaptación, nadie puede sobrevivir y ningún sistema puede funcionar. Sin embargo, a la inversa, también es cierto que casi todas las situaciones están tan abiertas a la interpretación que se utilizan puntos de vista o “perspectivas” muy disímiles de la interpretación del mundo, las cuales pueden persistir durante mucho tiempo.

La arquitectura mental incluye contextos culturales. Para mi suegra colombiana, Leticia Hernández de Duque (1923-2012), de Medellín, habría sido cultural y personalmente apropiado reaccionar ante la situación descrita con el dicho: “Los hombres no tienen nada que hacer

en mi cocina”. Era el caso en aquel entonces, pero hace mucho que eso cambió en Colombia. Y “Doña Lety”, como la llamaban todos, ¡supo adaptarse en sus últimos años! A diferencia de los programas digitales previamente conocidos, logró adaptarse bien a este caso especial de obsolescencia sociocultural en la vida cotidiana.

La clasificación de una situación o el encuadre determina los patrones de comportamiento o reacciones activados (afrentamiento). Ambos son contextual y situacionalmente diferentes, pero siguen contextos evaluativos de nivel superior que pueden describirse como un sistema de valores. La arquitectura mental de una persona no solo consiste en un sistema de valores; no obstante, resulta inconcebible sin actos de evaluación mental y tonos emocionales-afectivos.

Esto puede demostrarse, entre otras cosas, mediante el fenómeno del “disturbio ético” en su contexto histórico y sociocultural. Cuando nació mi primera hija, Sabrina, en 1980, compré un automóvil particularmente seguro, un Volvo usado que tenía más de 10 años. En aquel tiempo, esta marca sueca representaba el mejor estándar de seguridad posible. El vehículo aún no tenía cinturones de seguridad en el asiento trasero. Entonces, no podíamos abrochar a nuestra hija. Eso tampoco molestaba a nadie y, a menudo, pese a nuestras advertencias, ella jugaba feliz durante los viajes.

Si mi hijo intentara transportar a su hija, mi nieta Amalia, que nació en 2019, sin una silla infantil adecuada, de inmediato surgiría un “sentimiento de disturbio ético” muy masivo en todo el entorno familiar. El encuadre de la situación estaría asociado con “peligro para el niño” y “acción irresponsable”, pero también con las correspondientes evaluaciones emocionales. En consecuencia, habría reacciones al lidiar con la situación (afrentamiento) que claramente hablarían en contra de tales acciones.

Esos ejemplos se refieren tanto al contexto individual como al colectivo, típico del tiempo y sociocultural de las evaluaciones éticas. Si la arquitectura mental influye en nuestra evaluación ética y sistemática de situaciones –y, por ende, da forma a nuestro paisaje de valores individuales y colectivos–, este siempre debe entenderse como contextual y situacionalmente flexible.

Esto es fundamental para la ética digital pues lo dicho resulta en requisitos concretos para la ética de la programación y para las aplicaciones digitales.

Por un lado, debe asegurarse que los contextos y criterios relevantes de la evaluación, en el sentido del sistema de valores subyacente, se divulguen de modo transparente. Por otro lado, los programas deben ser “revisables” y tener una “fecha de caducidad” porque las premisas contextuales de su programación pueden cambiar. Esto no solo se muestra en los ejemplos del “hombre en la cocina” y los “cinturones de seguridad en el asiento trasero”, sino que incluye cambios en las evaluaciones que se hacen en la ley.

Tales cambios no pueden deducirse con precisión del horizonte de una argumentación utilitaria y deontológica de estándares normativos. Por esta razón, la contextualización y la capacidad de revisión de los sistemas de valores en una ética modernizada de la virtud y de la actitud deben verse como una ventaja sobre una validez general ficticia.

Sin embargo, esto se aplica únicamente en conexión con el estándar normativo de humanidad, ya que solo entonces están disponibles piedras de toque sustantivas, aunque controvertidamente discutidas. Tales piedras de toque tienen el propósito de prevenir un relativismo de valores y un deslizamiento hacia la barbarie, como en la dictadura nazi o en otras dictaduras.

La presentación del funcionamiento de los sistemas mentales y los valores asociados con ellos siempre es útil cuando se trata de argumentos éticos, especialmente la evaluación y priorización de decisiones y acciones.

No obstante, las aplicaciones digitales no pueden prescindir de tales evaluaciones y priorización. Debido a que las habilidades lingüísticas éticas son ahora parte de los requisitos para los profesionales digitales de una amplia variedad, la ética digital también debe incluir una breve visión de los métodos prácticos y reconocidos de toma de decisiones éticas.

Estos métodos son inicialmente independientes de las direcciones teóricas fundamentales utilizadas dentro de la ética.

El desafío de la responsabilidad: métodos de toma de decisiones éticas

No es de extrañar que en la ética, como en otras áreas de la vida, haya procedimientos profesionales para evaluar la acción. La ponderación de intereses se considera la solución ideal. Luego se toma una decisión de acuerdo con el mayor bien o el menor mal.

Sucede con regularidad que las opiniones divergen. Nos diferenciamos en nuestros modelos mundiales, planes de vida y arquitecturas mentales. Esto conduce a tensiones y conflictos, discusiones y malentendidos, una y otra vez. A veces, las emociones tienen un efecto más fuerte que los argumentos, lo que no es un buen requisito previo para una decisión éticamente responsable.

Una opción probada y comprobada para objetivar decisiones es la consideración ponderada de argumentos, que también se pueden representar de manera digital. Debido a que generalmente se juntan varios factores para una evaluación, el método de ponderar los intereses

puede diseñarse de tal forma que se asigne un valor en puntos para cada argumento a favor y en contra.

El valor total de los puntos otorgados no debe exceder un valor específico; por ejemplo, 100 puntos. En un evento interactivo, al pegar puntos en un rotafolio, por lo general se muestra rápidamente quién está ponderando qué argumentos de modo diferente y cómo, al final, se inclina la balanza.

Por su parte, es éticamente deseable la creación de transparencia sobre las razones de las diferentes evaluaciones en juego. Además, el método antes señalado –de ponderación de intereses con divulgación transparente de las ponderaciones realizadas– es muy adecuado para la operacionalización digital.

A veces, el mismo problema es visto por una persona como una ventaja y, por otra, como una desventaja. Si ambas tienen la misma cantidad de poder de decisión, habrá un empate en la toma de decisiones. Esto puede suceder sin ningún conocimiento digital si, en una pareja, una persona prefiere unas vacaciones junto al mar y la otra en la montaña. Si las dos mantienen alto el valor de la asociación, entonces pueden, por ejemplo, turnarse e ir al mar una vez y a las montañas la otra. Así se llega a la meta, a través del método de clarificación ética de valores, ya que ambas partes están de acuerdo en la manera de tomar decisiones ponderando un cierto valor como particularmente alto: el valor de la asociación.

Cuando se trata de programas digitales, es concebible que se pueda parametrizar la ponderación de los argumentos. Si esto no es posible, debe quedar claro para todos los usuarios el hecho de que el modo en que se lleva a cabo la ponderación también representa una forma de ejercicio del poder.

Otro método se relaciona específicamente con la ponderación de las consecuencias o del impacto de una decisión. En tal caso, se prescinde de juzgar si una decisión es “correcta” en el sentido ético. Quienes eligen el método de evaluación de impacto suelen estar en el proceso de preparar una decisión. Porque solo cuando las consecuencias se coordinan de antemano, según criterios adecuados, tiene sentido la ponderación de intereses.

Si, por ejemplo, unas vacaciones junto al mar solo son posibles en avión y unas vacaciones en la montaña lo son únicamente en tren, habrá parejas que valoren la sostenibilidad. Entonces, su evaluación de impacto tenderá a ser más en la dirección de “vacaciones en la montaña”, ya que esto permite una huella de CO² significativamente menor. Sobre la base de la aclaración de valor a favor de la “sostenibilidad”, la consecuencia de un consumo alto o bajo de CO² se considera relevante para la decisión. Sin embargo, esto no se aplicará a todas las parejas.

El problema con la evaluación de impacto radica en evaluar, de manera inequívoca, qué consecuencias se están considerando realmente y cómo se ponderan las mismas. Si las consecuencias primordiales se pasan por alto o se evalúan de modo poco realista, es fácil caer en la tentación de decisiones éticamente problemáticas.

En muchos casos, no obstante, el término “éticamente problemático”, o aun “poco ético”, se refiere a la falta de preparación suficiente para la decisión. Cualquiera que no se preocupe en la preparación razonable de una decisión importante siempre tomará una decisión poco ética, incluso si, en retrospectiva, esto resulta ser la decisión apropiada, por pura coincidencia.

Por lo tanto, la calidad del proceso de toma de decisiones pertenece a la ética. Si como una manzana o una pera durante un descanso del trabajo, puede no ser una decisión significativa, pero la construcción de

una nueva presa o de un nuevo aeropuerto es asunto de muchos y tiene una importancia política considerable. Entonces, se puede formular lo siguiente, como un criterio para medir la calidad de las decisiones éticas: “Cuanto más importante es una decisión y mayor su impacto, más profesionalmente debe prepararse y más cuidadosamente debe sopesarse”.

Además de la aclaración de valores y evaluaciones de impacto, el procedimiento de evaluación de riesgo-beneficio se incluye a menudo en la ponderación de intereses para una decisión. La evaluación financiera de una inversión, por ejemplo, es fundamental para las decisiones económicas. Luego, el beneficio y el riesgo se mapean principalmente en términos monetarios. Un plan de financiamiento asume que la liquidez utilizada inicialmente se ganará dentro de un cierto marco de tiempo y bajo supuestos realistas (“retorno de la inversión”). Lo mismo se aplica en el sector privado a la compra de una casa o un automóvil. Debe haber una probabilidad predominante de que, por ejemplo, un préstamo también pueda repagarse.

En no pocos casos, la evaluación de riesgos y beneficios orienta la acción. Sin embargo, es éticamente problemático si solo se considera una parte de aquellos como, por ejemplo, la mencionada dimensión monetaria.

Tomando esta idea, las empresas más grandes han creado sistemáticamente, en las últimas dos décadas, los llamados “mapas de riesgo”, mediante los cuales se evalúa la probabilidad de ocurrencia y la cantidad de daño de los posibles riesgos, desde hurtos en tiendas hasta grandes incendios, desde piratería hasta daños de reputación.

En el sector de los reaseguros, esta forma de análisis de riesgos llevó muy pronto al hecho de que las inversiones financieras en el sector del carbón se consideraban críticamente y no se realizan más. El trasfondo de tal decisión financiera no es solo el cambio social en la evaluación ética de los combustibles fósiles, sino la creciente difusión de los criterios

Si las consecuencias primordiales se pasan por alto o se evalúan de modo poco realista, es fácil caer en la tentación de decisiones éticamente problemáticas.



de inversión ESG en el mundo financiero. Se trata de la consideración de los riesgos ambientales (*environmental*), las consecuencias sociales (*social*) y la gestión empresarial conforme a las normas (*governance*).

No obstante, como suele ser el caso, no se puede llevar a cabo una evaluación de riesgo-beneficio sin la incertidumbre de las suposiciones realizadas. Esto también se aplica cuando las ayudas digitales están disponibles, por ejemplo, en el apogeo de la pandemia del coronavirus, en la primavera del 2020. Las reacciones demasiado tempranas tienen un impacto económico masivo, mientras que las demasiado tardías conducen a un número comparativamente mayor de muertes por contagio.

Por ende, la discusión sobre la calidad ética de una decisión debe tomar en cuenta la falibilidad de personas y máquinas. Si surgen nuevos contextos, se deben sacar nuevas conclusiones. Sin embargo, pese a toda falibilidad, una reacción bastante rápida a situaciones peligrosas es un sello distintivo de las personas: intentan adaptarse una y otra vez, cometen errores y los corrigen. Como resultado, en numerosos casos están dotadas de una “competencia en desastres” –propia de la especie humana– que es considerable (*cf. H. F. Angel 1996*) y que describe la compleja capacidad de sobrevivir y hacer frente a aquellos. Lo que no se quiere decir aquí es la capacidad humana para generar catástrofes y que se halla tan presente.

La “competencia en desastres” es difícil de representar digitalmente porque los desastres se caracterizan por lo inesperado y la secuencia de eventos que se prueba de antemano con una probabilidad muy diferente. Pero la programación que también cuenta con probabilidades aparentemente pequeñas en el sentido del “cisne negro” (N. N. Taleb 2008) corre el riesgo de volverse costosa y lenta. Demasiados escenarios tendrían que ser considerados pues, en este contexto, el término “cisne negro” significa la ocurrencia de eventos extremadamente improbables, ya que los cisnes negros son muy raros en la naturaleza. No obstante, si alguien desea tener en cuenta todas las eventualidades, un programa será costoso y engorroso.

Además, los programas de IA no son necesariamente mejores que el sentido común humano en caso de un desastre. Solo son mejores si no se trata de una catástrofe inesperada, sino de una secuencia desafiante de eventos que han sido considerados de antemano en el programa. Sin embargo, ello contrasta con las conclusiones que acabamos de indicar.

Resumiendo, el “maletín metodológico” o la caja de herramientas éticas para la toma de decisiones éticamente justificadas incluye, en particular, la ponderación de intereses sobre la base de una aclaración de valores éticos, la evaluación de impacto, la evaluación beneficio-riesgo y la reflexión de principios fundamentales.

La calidad ética de una decisión depende del nivel adecuado de preparación de la misma. Pero, dado que las personas solo tienen una vida útil limitada y las decisiones se toman en un tiempo razonable, es parte del profesionalismo ético evitar exageraciones en la demora o en el tiempo utilizado para tomar una decisión.

Tales exageraciones incluyen escrupulosidad y dilación, así como decisionismo e impulsividad falsa. Quien reflexiona durante tanto tiempo antes de decidir si comerse una manzana o una pera

porque se “muere” de hambre se convierte en víctima de sus propios escrúpulos o de su exagerada “holgazanería”, como antes se llamaba a la procrastinación. No obstante, aquellos que en pocos minutos toman decisiones sobre una nueva planta de energía basándose en su instinto parecen poco profesionales y pueden convertirse en víctimas de su adicción a la toma de decisiones y la impulsividad. La “medida apropiada” para preparar una decisión también depende de cómo se maneja el cronograma disponible, no solo de la dirección de la decisión que se ha tomado.

Volvamos a nuestro ejemplo: vacaciones junto al mar o en la montaña. Discutimos dos valores principales para una pareja casada: la asociación para la toma de decisiones y la sostenibilidad. ¿Pero qué cuenta más: la colaboración o la sostenibilidad? Básicamente, no hay una respuesta general a esta interrogante, pues para cada situación concreta de toma de decisiones existen numerosos parámetros que influyen.

Como se describió, las personas organizan su paisaje de valores intrínsecos o la “arquitectura de sus valores” (*cf. U. Hemel 2019, 335–350*) de modo jerárquico, a la manera de una pirámide de valores. Hay pocos valores superiores, varios valores destacados y numerosos valores base. Desde un punto de vista técnico o matemático, los valores pueden verse como vectores para la priorización fáctica de las acciones. En otras palabras, determinan la dirección preferida de nuestras acciones al ponderar diferentes valores situacionalmente importantes. El factor decisivo no es la retórica de los valores, sino la práctica del valor real: se trata de la priorización “fáctica”, no “reivindicada”, de las acciones humanas.

Esto conduce regularmente a tensiones entre el ideal y la realidad, entre un valor y el otro. Este “conflicto de valores” surge, por ejemplo, cuando la rapidez y el cuidado no se pueden lograr juntos: un gran cuidado suele llevar más tiempo, pero falta cuando las decisiones se

toman a gran velocidad. Por lo tanto, son necesarios los “compromisos en ponderar los valores éticos” que conduzcan a acciones concretas.

Podemos concluir que cada acción humana es un mapeo matemático. Sin embargo, debido al carácter de caja negra de la práctica humana, no resulta fácilmente determinable entre las diferentes opciones de acción, que representan –consciente o inconscientemente– valores diferentes y, al mismo tiempo, el conjunto de nuestra “arquitectura mental”.

Cada acto u omisión real es, por así decirlo, un acto de equilibrio, el buscar y encontrar un curso de acción óptimo que refleje los valores fundamentalmente requeridos, pero también apropiados para la situación, de una persona o grupo de personas de la mejor forma posible.

Reflexión sobre principios: la responsabilidad digital en su sentido moral y legal

Los métodos de toma de decisiones éticas descritos brevemente, en particular el proceso de clarificación de valores, deben complementarse con una reflexión sobre principios en este contexto de posibles conflictos de valores éticos. Dado que los principios guían la interpretación de los valores, son los hitos de la orientación a la acción.

La disposición de los valores en una jerarquía de valores no es fija; depende de las experiencias de aprendizaje, los desafíos situacionales y los contextos sociales. No “vemos” los valores anclados en nosotros, y solo podemos inferir los valores de los demás a partir de su comportamiento.

Esto requiere actos de interpretación que, a su vez, pueden matizarse subjetivamente. La cuestión de comprender la acción humana y la relación entre acción e interpretación es tan emocionante como compleja. Es el tema de la reflexión filosófica, sociológica y teológica, aunque no se tratará aquí con más profundidad (*cf. H. J. Gadamer 1960, K. Demmer 1985, W. Edelstein, M. Keller 1982, K. Joisten 2009*).

Pero si las personas ya son una “caja negra” para los demás, y a veces incluso para ellas mismas, en términos de sus valores, entonces la atribución de ciertos valores a un comportamiento observado no es tan clara como desearíamos. Después de todo, estamos de acuerdo con que las acciones humanas se pueden atribuir a una persona específica que es responsable.

En el mundo digital, la relación entre acción e interpretación, resultado y comportamiento, programación y responsabilidad, plantea nuevas y adicionales preguntas. En aplicaciones complejas, la salida de una máquina programada digitalmente podría interpretarse como “comportamiento” y, por ende, como el primer paso en la atribución de responsabilidad. Luego llegamos al área de la ética de la máquina y la responsabilidad de la máquina. Aquí debemos preguntarnos si existe un comportamiento sin responsabilidad. La respuesta es “no” cuando relacionamos el comportamiento solo con los humanos. Diferente es el caso cuando, además de los humanos, atribuimos a los animales y a las máquinas no solo la producción, sino el “comportamiento” responsable.

La biología del comportamiento (*cf. K. Lorenz 1978*) analiza el comportamiento de los animales sobre la base de una observación intensiva, sin aplicar categorías éticas. Si este pensamiento se trasladara a las máquinas, en la ética digital –análoga a la etología cognitiva de los animales– nos encontraríamos ante un nuevo tipo de “etología cognitiva de las máquinas”. Lo atractivo de esto es que la cuestión de las evaluaciones éticas del comportamiento de la máquina puede excluirse metódicamente. La atención se centra en el resultado, el comportamiento; no en la evaluación del comportamiento de las máquinas.

Una reflexión profunda sobre los principios éticos argumentará que la etología cognitiva de las máquinas es, al menos, insuficiente. Los animales están determinados, en gran medida, por su ADN y otras características. Los seres humanos pueden tener la responsabilidad del

bienestar animal y de su trato con los animales, también con miras a las intervenciones en la línea germinal y los clones. No obstante, los humanos no son necesariamente responsables del comportamiento de los animales, pero hay excepciones, por ejemplo, con las mascotas. De cierta manera, los perros reflejan el enfoque educativo de sus dueños, por lo que estos son indirectamente responsables del comportamiento de sus animales.

La ética digital, que se relaciona con la producción de aplicaciones y máquinas programadas digitalmente, está más cerca de las mascotas que de los animales salvajes cuando se considera la atribución de responsabilidad. Nos guste o no, los productos digitales –como robots, androides o software basado en IA– son una creación humana a la que todavía no le hemos atribuido su propio libre albedrío.

Volvamos a la pregunta sobre “comportamiento sin responsabilidad”. Si la respuesta a la interrogante en torno a la existencia de un comportamiento sin responsabilidad es “no”, volvemos al campo de la acción humana. La salida de las máquinas sería, entonces, una extensión, un “banco de trabajo extendido”, un instrumento exteriorizado de la acción humana. En consecuencia, las personas que crearon las máquinas también son responsables del comportamiento de estas.

Con esta línea de pensamiento se podría justificar una forma de hablar que se refiere a “ética anclada en las máquinas” y de “paisaje de valor de las máquinas”. Solo los sujetos humanos –portadores o programadores de tales valores de máquinas– son, en el mejor de los casos, invisibles; en el caso más difícil, no pueden ser identificados en absoluto o, incluso, han muerto.

Por consiguiente, la pregunta sobre el “lugar” de la responsabilidad del comportamiento de la máquina no es fácil de contestar. La exigencia de responsabilidad (*H. Nissenbaum 1994, 73-80*) en la

programación de la inteligencia artificial va exactamente en la dirección de estas interrogantes. Dado que la responsabilidad también conlleva responsabilidad legal, la reflexión sobre los principios de la ética digital nos enfrenta a un dilema previamente no resuelto.

El problema de la responsabilidad digital incluye el conocimiento de que prácticamente no existe un software libre de errores. Aun con una probabilidad de 1:1.000.000 –es decir, con un error por millón de líneas de programación (líneas de código)–, habrá decenas de errores en softwares complejos con varios millones de líneas. Solo en el nuevo VW ID3, el software de la unidad de control para coordinar las distintas unidades de control requiere 20 millones de líneas de código de software, como afirmó el proveedor, el CEO de Continental, en una entrevista (*Schwäbische Zeitung*, 13 de junio del 2020, p. 7). Los errores de software no suelen ser dramáticos, pero sí nos señalan los límites de la responsabilidad.

Al fin y al cabo, es casi imposible rastrear al programador individual en busca de un error de programación. Aparte de eso, podría haber dejado la empresa hace mucho tiempo o, como se mencionó antes, haber fallecido. La compañía responsable tampoco existe para siempre; incluso las muy grandes pueden hundirse. La solución podría encontrarse donde haya un consenso en la arena política para establecer una especie de fondo de la industria para las fallas digitales. Ya existe algo similar con el seguro concursal para operadores de viajes.

El problema de fondo va más allá de la protección social y legal contra los riesgos de las aplicaciones digitales. Básicamente se trata del problema de la difusión de responsabilidad. Considerando una reflexión sobre los principios, también se trata de ver difuminada la frontera entre las acciones humanas y la responsabilidad por las aplicaciones digitales.

Las máquinas y los programas deben configurarse o “parametrizarse”. Por lo tanto, la salida digital puede verse como resultado de la programación anterior y la configuración realizada por el operador. Pensemos nuevamente en el ejemplo del robot de cuidado que registra e informa el consumo de chocolate de un residente mayor. ¿La responsabilidad de dicho comportamiento de notificación recae en el fabricante o en el operador? Estas y otras cuestiones similares serán decididas por los tribunales en el contexto social a su debido tiempo.

Sin embargo, la difusión de la responsabilidad en un mundo híbrido tiene otra cara –que podemos describir con los términos “pseudopersonalidad” y “pseudosocialidad”– en determinadas aplicaciones digitales. La peculiaridad de las pseudopersonas es que el compañero de interacción humano sabe cognitivamente que su contraparte no es una persona humana, aunque no se siente así emocionalmente. Es decir, la otra persona parece una persona real. Entonces, se trata de una cierta forma de disonancia cognitivo-emocional.

“Siri” y “Alexa” le hablan al usuario, pero no son humanos. Vale la pena analizar por separado si debemos considerarlos como pseudopersonas o cuasipersonas. En el caso de las cuasipersonas, la diferencia entre hombre y máquina es aún más borrosa que en el caso de la disonancia cognitivo-emocional, la que al menos se puede hacer consciente fácilmente.

Los robots pueden simular o emular emociones. Esto va tan lejos que permiten, inclusive, que el sexo digital se experimente como algo placentero (cf. https://www.wuv.de/tech/smart_sex_das_liebesleben_der_zukunft, consultado el 12 de junio del 2020, a las 4:29 p.m.). La imaginación humana complementa lo que falta en la contraparte humana real. Las llamadas bodas en solitario o sologamia son una forma extrema: una persona se casa con sí misma (cf. <https://en.wikipedia.org/wiki/Sologamy>, consultado el 12 de junio del 2020, a las 4:36 p.m.). En Kioto (Japón), por

ejemplo, hay una agencia de viajes especializada en esta práctica (*ibid.*). No sabemos hasta qué punto las parejas sexuales digitales con gafas VAR desempeñan un papel en entornos tan solitarios. Ello se puede dejar al caso individual.

Hasta ahora, estos han sido casos extremos, pero la tendencia a humanizar una contraparte digital aumentará en el futuro, no solo en el sector del cuidado, en el mundo del juego o en el área del placer sexual. Los robots, los androides y los dispositivos digitales no son humanos; no obstante, pueden crear la ilusión de una contraparte humana mediante la salida de voz, el diseño de la superficie y un repertorio de comportamiento apropiadamente programado.

La responsabilidad también se puede difundir en esta área, como en el caso del comportamiento adictivo digital (*cf. U. Braun 2014*). Desde el 2018, por ejemplo, la adicción al juego en línea se ha incluido como una adicción no material en el catálogo de diagnóstico de la Organización Mundial de la Salud, según la CIE 11 (6 C 51.0).

Los avatares pueden entenderse como pseudopersonas. En su mayoría, son representantes digitales o alias del jugador. Dado que la persona es libre de elegir las propiedades de sus avatares, puede modificar las propiedades naturales normalmente inalterables, como la edad, el sexo o el color de la piel. Desde ese punto de vista, los avatares son pseudopersonas porque actúan como personas, pero no lo son.

Al final, la responsabilidad digital surge de la interacción entre las personas y la sociedad. Nadie por sí solo determina qué significa justicia, quién tiene acceso a determinadas plataformas de internet, a qué edad y quién es responsable de la mala conducta. Sin embargo, está claro que los procesos de educación digital y ética de los medios digitales (*cf. G. Ulshöfer, M. Wilhelm 2020*) plantean enormes desafíos para los padres y las familias, pero también para las escuelas y los

proveedores de formación profesional. El objetivo es esa interacción de la razón y la emoción que nos permite asumir la responsabilidad de nuestra propia vida.

De ningún modo está clara, además, la decisión sobre qué valores o principios vemos anclados en una situación específica. Nuestra interpretación situacional del desempeño de los valores puede variar ampliamente. Cuando estamos en una ciudad calurosa con amigos, es más probable que ver una buena heladería actualice el valor “entusiasmo por la vida a través del disfrute del helado” que activar el valor “entrenamiento de disciplina mental a través de la renuncia”.

Pero, ¿cómo se debe configurar una aplicación para motivarnos a llevar una vida saludable? Al final de cuentas, puede ser que simplemente estemos llenos por el desayuno y no tengamos ganas de tomar un helado. O tal vez solo tenemos miedo de manchar nuestra ropa. Entonces, la decisión “en contra de un helado” sería, en el mejor de los casos, una expresión del valor de no hacerse daño a uno mismo. Sea como fuere, la ambigüedad constitutiva de las situaciones de la vida no se puede mapear de manera completamente digital.

Sin embargo, la falta de claridad en la atribución de valores al comportamiento externo es solo una cara de la moneda, porque la huella de cada acción concreta, cuando se repite, conduce a estructuras mentales estables, es decir, a “actitudes”, detrás de las cuales hay valores interiorizados de individuos u organizaciones.

Como ya se mencionó, la totalidad de nuestras actitudes éticas representa nuestro “paisaje” o conjunto personal de valores, cuya estructura, idealmente, sigue unos pocos principios comprensibles, entre ellos la reciprocidad (como se formula en la Regla de Oro) o la búsqueda de un equilibrio entre el interés propio y el bien común. Esto

también se refleja en el doble mandamiento bíblico del amor (“Ama a tu prójimo como a ti mismo”).

No obstante, el equilibrio entre el interés propio y el bien común va más allá de la acción individual, y busca estructuras y procesos adecuados en contextos más amplios. Por lo general, el término “clima” se definiría como un bien común, por lo cual habría que debatir cuestiones de responsabilidad ecológica, que también se refiere a la huella de carbono y su costo por efecto de las aplicaciones digitales, como la transmisión por secuencias (*streaming*).

Debido a que no siempre podemos tener una discusión fundamental en la rápida sucesión de situaciones de la vida, y a que nuestros valores no son libremente intercambiables, necesitamos la estabilidad de este paisaje de valores habitual. A su vez, ello no es independiente de nuestro contexto sociocultural. Se superpone con el sistema de valores de otras personas, aun cuando es individual en sus sutilezas. Se forma a través de acciones repetidas con preferencias de valor similares.

Las acciones repetidas aumentan la probabilidad de que se elija la misma opción en situaciones similares. Es lo que se quiere decir con la creencia de que las acciones conducen a actitudes. Esta “habitualización” se aplica a nivel individual y a nivel colectivo, por ejemplo, en el ámbito económico con ética empresarial “habitual” (*cf. U. Hemel, A. Fritzsche, J. Manemann 2012*).

En resumen, nuestras acciones plasman actitudes; las actitudes se transforman en sistemas de valores; los sistemas de valores moldean rasgos de caracteres y, por fin, “virtudes”. Entonces, en este punto, volvemos a la “ética de la virtud”.

En la práctica, significa que no podemos predecir con absoluta certeza, ni para nosotros ni para los demás, cómo nos comportaremos

en una situación determinada. Sin embargo, el “paisaje de valores” relativamente estable formado en una persona conduce a hipótesis muy probables. “Como lo conozco, no pasa por una heladería sin pedir un helado”, sería una declaración típica sobre un amigo en la escena descrita anteriormente.

La interpretación de los valores y su relación con los principios no es arbitraria en nuestra experiencia subjetiva. Hay algunos valores humanos máximos a los que seguimos volviendo individual y colectivamente. Estos incluyen la verdad y la veracidad, la equidad y la justicia, la no violencia y la amistad, pero también la asociación y la sostenibilidad. No es una coincidencia que, en octubre del 2018, el Parlamento de las Religiones del Mundo afirmara solemnemente dichos valores como valores compartidos por todas las religiones representadas (cf. *K. J. Kuschel 2019, 39-54, S. Schlensog 2019, 28-38*).

En contraste con tales valores, las virtudes cardinales clásicas están más orientadas hacia los rasgos de carácter individual. En el lenguaje actual podrían verse como competencias éticas clave. Estas son las virtudes de la prudencia (*prudentia*), la valentía (*fortitudo*), la justicia (*iustitia*) y la templanza (*temperantia*).

Desde la perspectiva actual, en el caso de la *temperantia*, buscaríamos una conexión con la *prudentia*, y probablemente hablaríamos más de juicio inteligente, compromiso valiente, justicia, equilibrio o sentido del equilibrio.

En la ética, incluida la ética digital, no se trata solo del comportamiento de las personas en el sentido de la ética individual, sino también del comportamiento de las instituciones, organizaciones y empresas en el sentido de la ética social.

Como muestran los ejemplos descritos, el equilibrio entre la ética individual y la ética social se enfrenta a nuevos desafíos, especialmente en el mundo digital. Con bastante frecuencia, los intereses de una institución, una corporación o un Estado son órdenes de magnitud más poderosos que los del individuo. Por ende, una reflexión ética sobre los principios requiere de algunos principios socioéticos rectores que vayan más allá de las virtudes individuales.

Precisamente por eso merece la pena echar un vistazo a los principios básicos de la doctrina social cristiana. Se trata de la tríada de “personalidad, subsidiariedad y solidaridad” con miras al bien común. El valor central es la prioridad de la persona en el sentido de personalidad, porque los sistemas económicos y sociales no son un fin en sí mismos. Tienen que servir al ser humano y, ante todo, a la persona individual (*cf.* U. Hemel 2013).

Dado que las decisiones se pueden tomar de forma centralizada o descentralizada en las organizaciones sociales, el segundo principio básico de la ética social cristiana es la subsidiariedad. Las decisiones deben tomarse al nivel directamente afectado, en la medida de lo posible. Es el lugar correcto de la responsabilidad.

No obstante, los intereses del individuo o de los departamentos descentralizados no deben ser a expensas del público en general. Por consiguiente, el tercer principio, la “solidaridad”, se refiere al nivel adecuado de cohesión. El bienestar de la persona y el bien común deben equilibrarse.

Tal reflexión sobre los principios ético-sociales tiene sus ventajas de cara a la ética digital pues, al contrario de numerosas publicaciones, no se trata solo de la ética individual de los programadores, directores generales u otras personas responsables o, a la inversa, de los consumidores digitales, de los *clickworkers* u otros usuarios. Siempre se trata de los efectos en la vida social en su conjunto.

Observar a ambos lados, la ética individual y la ética social, agudiza la conciencia de posibles conflictos de intereses y constelaciones de poder. La realización de valores nunca ocurre en un espacio libre de poder. Está conformada, fundamentalmente, por un contexto de intereses y opciones de acción que se basan en el poder efectivo o en el poder del argumento.

Esto hace que la ética –incluida la ética digital– sea tan difícil como necesaria: debe ofrecer principios más allá de la constelación del poder y ampliar el espacio para la argumentación. Pero también debe ser capaz de lidiar con interpretaciones de valores en el ámbito del poder.

Si el principio básico de humanidad se cita aquí constantemente por una buena razón, entonces, una tarea de la aclaración de los principios siempre incluye la reflexión sobre la situación respectiva y la constelación que está en juego. Porque de aquellas situaciones y constelaciones surgen perspectivas del mundo y la visión de valores que pueden ser muy diferentes, aun contradictorios.

La industria automotriz, por ejemplo, estimará el valor de la movilidad en la práctica de modo diferente a una asociación ambiental. Los valores no se exigen ni se pasan por alto en una zona libre de conflictos. Este es un caso de cómo, en su forma lingüística, los “valores” siempre pueden ser objeto de intereses retóricos, incluyendo la “retórica de combate ético”.

Se puede hablar de esta retórica cuando los valores se instrumentalizan para los propios objetivos y propósitos. Porque hablar de valores tiene no solo un carácter descriptivo, sino también publicitario: es una forma especial de autorretrato que habla de un ideal que a veces es mejor, y a veces peor, cubierto por la realidad.

Ética en el mundo digital y ética profesional digital

En el mundo digital no se aplican valores diferentes a los de otros entornos de vida. Por ello, la reflexión anterior sobre principios en el equilibrio entre la ética individual y la ética social se refiere tanto al mundo analógico como al digital.

El concepto de mundo de vida (*Lebenswelt*) utilizado en esta sección fue introducido en 1932 por Alfred Schütz (1899-1959) para describir el conjunto de huellas y patrones interpretativos de la vida cotidiana que normalmente no se reflejan de manera específica (cf. *A. Schütz 1974*). Por su parte, Alfred Schütz estuvo fuertemente influenciado por los análisis fenomenológicos de Edmund Husserl (1859-1938), lo que enfatizó la diferencia –de hecho, la distancia– entre el mundo de la ciencia y la vida cotidiana (cf. *E. Ströker 1979*).

Precisamente porque vivimos en una época que se caracteriza por el paralelismo de mundos profesionales y privados muy divergentes en la ciencia, los negocios, la educación y la política, una “crítica de la razón digital” debe mirar las interrelaciones entre estas áreas. Alfred Schütz es pionero aquí: fue el primero en abordar la relación entre el mundo y las “formas de vida” en un lenguaje que todavía es sorprendentemente fresco en la actualidad (cf. *A. Schütz 1981*).

El concepto del mundo de vida también puede ser útil en la transformación digital, ya que esta se caracteriza por el hecho de que no solo genera cambios técnicos, sino que tiene consecuencias sociales, hasta la convulsión de la vida cotidiana y del mundo del trabajo.

La pandemia causada por el coronavirus, que estalló en enero de 2020, tuvo el efecto de un impulso adicional de digitalización en muchos países. La educación abierta y la cultura abierta se han convertido en palabras clave para acelerar el acceso digital a la educación y a las ofertas culturales digitales. Los cursos universitarios completos solo se

llevaron a cabo en formato digital; las conferencias de liderazgo y las rondas ministeriales, con videoconferencias.

La distinción entre ética individual y ética social se aplica, además, a los distintos mundos digitales. En el ámbito individual, se trata de nuestro manejo personal de los datos, pero también del alcance de nuestro autocontrol en relación con el potencial de dependencia digital y adicción a las redes sociales. Se trata de los derechos de libertad digital frente al Estado y la sociedad: ¿Cuánta privacidad tengo? ¿Debo estar resignado o indiferente al rastreo de las huellas de mis datos casi universal, hasta el último detalle, como algo inevitable en los tiempos digitales, o hay formas y medios de comportarme de manera diferente?

La crisis del coronavirus ha llevado las preguntas a un nuevo nivel, por ejemplo, cuando la palabra clave “rastreo” significa rastrear los datos móviles de las personas infectadas en interés de la salud pública. Dado que la UE es sensible a las cuestiones de libertad y protección de datos, se aplica el concepto arquitectónico de “rastreo paneuropeo de proximidad con preservación de la privacidad”, que observa el cumplimiento de las normas de protección de datos europeas y alemanas, y accede, de modo anónimo, a los contactos epidemiológicamente relevantes de las últimas tres semanas, sin registrar el perfil de actividad del usuario. La aplicación de advertencia del coronavirus del Gobierno federal alemán –a la que se puede acceder desde junio del 2020– se ha desarrollado respetando las normas arriba mencionadas.

De cualquier manera, el trabajo y la profesión siguen formando parte del mundo digital. Y aquí surgen viejas preguntas en una nueva figura. Si el teletrabajo se vuelve regla, en lugar de excepción, ¿qué pasa con la seguridad de datos necesaria? ¿Hasta dónde llega el control digital de los tiempos de trabajo y descanso? ¿Qué significa el liderazgo en un mundo digital en el que incluso la actual canciller Merkel tuvo

que trabajar desde casa a fines de marzo del 2020 debido al contacto con un médico que pudo estar infectado con COVID-19?

A su vez, ¿cómo afecta la transformación digital a la educación y las condiciones sociales? ¿Cuáles son las consecuencias para la cuestión de la inclusión y la exclusión, del acceso y la participación? ¿En qué medida la sociedad civil local, regional, nacional y mundial es responsable de permitir dicha participación a través del acceso digital?

Más aún, ¿qué conceptos son adecuados para transmitir la competencia digital y la educación digital de tal modo que las posibilidades de internet conduzcan a un desarrollo más amplio, en vez de encerrar al individuo en su propia burbuja de filtro? ¿Cómo logra combinar el uso de los medios digitales con actitudes éticas que promuevan la autodeterminación y las habilidades para la vida, en lugar del control externo y la dependencia? Entonces, ¿cómo debería diseñarse una ética de los medios digitales? (*cf. J. Bedford-Strohm, A. Filipovic 2020, 47-63*).

Estas preguntas tienen alcances que van hasta la ética de los deportes digitales (E-Sport) y la ética de los juegos digitales (Gaming). De tal manera, impactan profundamente en nuestro mundo social cotidiano; llegan al mundo de los trabajadores de los medios de comunicación y los trabajadores sociales; a los que participan en la educación escolar y a los teólogos, los activistas y las familias individuales. Surgen tanto entre semana como los domingos y días festivos.

Sin embargo, es justamente esta penetración digital de las diversas formas de la vida cotidiana lo que conduce a la necesidad de que no solo desarrollemos una ética del mundo digital, sino también, paso a paso, una ética profesional digital. Esta se centraría en la amplia gama de campos de aplicación digitales, con sus preguntas y contextos específicos.



La educación abierta y la cultura abierta se han convertido en palabras clave para acelerar el acceso digital a la educación y a las ofertas culturales digitales.

En el mundo laboral, solo una ética profesional digital nos prepara para las características particulares del mundo digital con sus elementos de continuidad y discontinuidad. Hoy en día, la competencia profesional incluye las competencias digitales específicas de cada profesión. Es una exigencia central, no obstante, entender que tal competencia digital deberá comprender, asimismo, unas habilidades lingüísticas y de acción éticas.

La ética en el mundo digital se enfrenta, por lo tanto, a la tarea de repensar todos los ámbitos sociales y personales de la vida. Y esto se aplica, en particular, a la cuestión de quién tiene qué responsabilidad en el espacio digital. Por tal motivo, siempre hay que hacer la pregunta: ¿En qué medida podemos y debemos asignarle la responsabilidad de sus acciones al individuo en una sociedad altamente individualizada? ¿Hasta dónde se extiende la responsabilidad de las organizaciones?

Y si es cierto que toda forma de tecnología ya contiene decisiones preliminares sobre la responsabilidad y la difusión o desdibujamiento de la responsabilidad, ¿no debería ser posible prestar atención a las cuestiones éticas a la hora de programar? Este es el enfoque que persigue la exigencia de ética por diseño (*Ethics by Design*), es decir, la consideración de la ética en la planificación e interpretación, el “diseño” de productos digitales.

La ética como parte de la tecnología digital: ética por diseño y diseño sensible a valores éticos

Un martillo se puede utilizar para fijar clavos, pero también como arma homicida. Lo mismo se aplica a un automóvil, que está sujeto a dicha ambivalencia de uso funcional y posible abuso. En la cultura científica y social actual, la exigencia de la “neutralidad” ética de la tecnología se deriva del uso éticamente positivo de un lado y del uso inadecuado de la tecnología del otro.

Esta idea, sin embargo, ha sido criticada con razón. Por supuesto, es correcto que las centrales nucleares se empleen para generar electricidad sin CO², aunque se pueden utilizar, además, para el enriquecimiento militar de uranio. También es cierto que la posibilidad de uso indebido va de la mano con un esfuerzo de seguridad mucho mayor, porque las centrales nucleares podrían ser susceptibles de sabotaje en determinadas circunstancias. Mientras tanto, la cuestión del almacenamiento de las barras de combustible enriquecidas no se ha resuelto, por lo que existe un gran escepticismo sobre las plantas de energía nuclear en muchos países. Como resultado, se inició la búsqueda de formas regenerativas y más descentralizadas de generación de energía.

Este ejemplo muestra que cada producto técnico y cada tecnología están asociados a implicaciones éticas directamente vinculadas con esa tecnología y que conducen a evaluaciones controvertidas. La tecnología, por ende, no es “neutral” sino portadora de “valores incorporados”, los cuales pueden discutirse de manera controvertida y ponderarse de modo diferente. Pero esta realidad no quita el hecho de que los valores son parte de todo tipo de tecnología.

La demanda de ética por diseño toma en serio tales consideraciones, y quiere situar la reflexión ética sobre los beneficios y los límites éticos de un programa en el inicio de la planificación y el diseño.

Hay cinco principios que deben observarse:

- Responsabilidad (*Accountability*)
- Alineación con los valores de los grupos de usuarios (*Alignment*)
- Explicabilidad (*Explainability*)
- Cumplimiento de los derechos sobre los datos de los usuarios (*User Data Rights*)
- Justicia (*Fairness*)

Con estas barandillas conceptuales, la ética por diseño crea su propio mundo lingüístico que promueve la reflexión ética y la capacidad de discurso. Esto incluye variaciones de los valores fundamentales requeridos, por ejemplo, mediante la adición explícita de valores como transparencia (*transparency*), sostenibilidad (*environmental sustainability*) y confiabilidad (*reliability*) o mediante matices más lingüísticos como, por ejemplo, cuando se trata de justicia en vez de equidad o cuando se habla del derecho a la privacidad (*privacy*) en lugar del respeto de los derechos de los datos (*User Data Rights*), según un documento publicado en abril del 2020 por un grupo de estudio en nombre de la VDE y la Fundación Bertelsmann (<https://www.ai-ethics-impact.org/en>, consultado el 14 de julio del 2020, a las 9:01 p.m.).

Esto no cambia el hecho de que los mundos del discurso de los programadores, filósofos de la tecnología, éticos y profesionales digitales no son idénticos a los intereses y el “mundo de vida” de los usuarios y un público más amplio. Por consiguiente, siempre queda imperfecto el esfuerzo por evitar un “sesgo” (*bias*), es decir, una distorsión o una unilateralidad de la propia perspectiva.

No obstante, la ética desde el diseño y el diseño sensible al valor (cf. *B. Friedman, D. G. Hendry 2019*) son pasos importantes en la dirección de la “innovación responsable”, donde los efectos éticos y sociales ya están involucrados en la investigación y creación de nuevos programas, con o sin inteligencia artificial.

El término “diseño sensible al valor” proviene originalmente del campo de la interacción cerebro-computadora, pero también se usó para áreas como identidad, tecnología de sensores o nanotecnología en el sector farmacéutico (cf. *F. Dechesne et al. 2013, 173-181; J. Timmermans y otros 2013*).

Básicamente, parece haber una creciente consideración sobre las consecuencias éticas y sociales de la tecnología digital (cf. *P. Kirchschräger 2020*). No todo lo que es posible tiene sentido ético y social. Definir dónde están los límites es una cuestión de debates sociales y, más allá, una cuestión de política.

En una comunidad democrática, sin embargo, los debates sociales y la política requieren de una amplia participación. Puede suceder que los caminos incorrectos solo se reconozcan después de un lapso y que las buenas soluciones necesiten tiempo para implementarse.

En definitiva, todavía estamos en el inicio de la transformación digital del mundo y la sociedad. Por lo tanto, en mi opinión, las cuestiones sobre responsabilidad y equidad en el uso de la tecnología digital serán aún más importantes en los próximos años (cf. *H. Nissenbaum 1994, 73-80*).

La equidad practicada y percibida por los usuarios genera confianza (cf. *N. Copray 2010*). Pero la confianza es el recurso crucial para la aceptación de nuevas tecnologías que realmente deberían beneficiar a las personas.

Debido a que la equidad (*fairness*) ha pasado a primer plano como un valor superior del diseño social—más allá de las fronteras nacionales—, en la siguiente sección abordaré la cuestión de la equidad digital como un nuevo desafío, especialmente para las empresas y las organizaciones.

La equidad digital como tarea de empresas e instituciones

Cuando se trata de la arquitectura mental de organizaciones e instituciones, se puede hablar de sistemas de valores, visiones del mundo y “arquitectura mental”, como ocurre con los individuos. En el caso de los usuarios humanos, por lo general no se reflejan explícitamente ni se plantean para la discusión, sino que se hacen notar mediante un cierto tono emocional de su propia experiencia de usuario.

Un cliente puede sentirse como un “factor de disturbios y molestias” porque tiene que hacer clic laboriosamente en menús engorrosos. Pero, además, puede disfrutar de la agradable sensación de recibir una ayuda significativa. En el mundo empresarial, se trata del término experiencia de usuario o “user experience” (UX). Si la orientación al valor se toma en serio para las empresas digitalmente activas, la equidad debe reflejarse en la experiencia del usuario.

Hasta la fecha, esa tarea solo ha sido reconocida parcialmente. Y aquí, también, la experiencia del usuario debe ir precedida de decisiones específicas de la empresa. En la práctica, no obstante, el aspecto ético del comportamiento de las empresas digitales aún parece en gran parte desinformado y sin esfuerzos de reflexión consciente. Por ende, vale la pena examinar más de cerca el término “equidad digital” (*cf. U. Hemel 2016*).

Veo, al menos, siete requisitos parciales para la equidad digital (*digital fairness*):

- Autodeterminación informativa (*Privacy by Design*)
- Reciprocidad digital
- Transparencia
- Trazabilidad (*Tracing und Tracking*)
- Fecha de caducidad regulada
- Escalación regulada
- Credibilidad digital

La “autodeterminación informativa” se vincula al valor de la libertad, que se expresa en la autodeterminación de una persona. Para las empresas digitales, implica la observancia del derecho a la privacidad, por ejemplo, a través de disposiciones para la privacidad desde el diseño. Esto significa que las aplicaciones y los programas están diseñados de tal modo que los usuarios deben decidir activamente ser “generosos” con la transferencia de sus datos. De lo contrario, se aplica un ajuste de fábrica bastante “restrictivo”.

La idea básica de la autodeterminación informativa es el derecho a los propios datos. Por consiguiente, quien genera datos debe tener derecho a decidir cómo se utilizan.

Tomemos nuestros datos de movimiento como ejemplo. Aquí espero de un fabricante de teléfonos inteligentes y de un fabricante de automóviles que yo tenga fácil acceso a mis propios datos, al menos en forma de datos sin procesar. Esto no descarta oportunidades para proveedores comerciales; por ejemplo, si me ofrecen una tarifa para ciertos servicios de datos o cuando me ofrecen un precio para la

respuesta a mi pregunta personal: “¿Dónde estaba yo exactamente el 9 de octubre del 2019?”.

La reciprocidad digital parece ser difícil de representar en la relación entre un solo usuario y grandes empresas de datos. Sin embargo, el valor de la reciprocidad ya se establece en la “regla de oro”: no debes tratar a los demás de manera diferente a como quieres que te traten a ti. Así que es un “dar y recibir”.

Esto podría incluir la siguiente regla: si un usuario genera datos de forma voluntaria, recibe acceso a determinados análisis o remuneración por los datos suministrados. La idea de una “nube pública” y una “sociedad fiduciaria de recopilación de datos (RD)”, desarrollada en el Capítulo 5, va en esa dirección y promueve los inicios de una “democracia de datos”.

La reciprocidad digital fomenta la participación. Por lo tanto, tiene un valor emocional inmediato cuya importancia no debe subestimarse. Los “servicios de equidad” o los “derechos de acceso” podrían convertirse en una nueva moneda para proporcionar y generar datos digitales. Esto es particularmente cierto cuando se brindan “servicios de valor agregado”, como el acceso a evaluaciones o comparaciones agregadas. Ya existen prácticas que propenden a ello y que mejoran la “experiencia de usuario”.

La transparencia es uno de los valores centrales de la sociedad civil actual. Queremos saber quién está detrás de una empresa, qué sucede con nuestros datos y cómo deben utilizarse. Detrás del valor de la transparencia se halla, entre otras cosas, el valor de la honestidad y la confiabilidad, pero también la información sobre posibles oportunidades y riesgos de la generación y uso de datos. Lo contrario, la falta de transparencia, significa pérdida de control, descontrol y posibilidad de prácticas desleales. La transparencia es un activo muy

valioso, más aún en el mundo digital, y está estrechamente relacionada con la “justicia percibida”.

La trazabilidad es un desarrollo posterior típico que surge del mundo digital. La posibilidad técnica de “rastreo y seguimiento” ofrece nuevas posibilidades de control, de manera que la “rastreadabilidad” se relaciona con el área privada y la discusión en torno a la cuestión de la adecuada protección de la privacidad.

En este sentido, los intereses de los empleados y los clientes digitales son dignos de protección. Veamos un ejemplo: si un empleador puede realizar un seguimiento, no solo de los viajes de negocios, sino de los viajes privados de un representante de ventas, ello iría demasiado lejos según nuestra comprensión de la libertad. En el 2020 se discutió intensamente si el Estado debería poder utilizar perfiles de movimiento para controlar la propagación del coronavirus y, en el contexto democrático en Europa, llevó a un consenso predominante: sí, eso es aceptable, pero solo con el consentimiento de los afectados.

El requisito de una fecha de vencimiento regulada para el uso de los datos generados es otra característica específica del mundo digital. En el ámbito político, el tema se debatió bajo la cuestión de la “retención de datos”, aunque también se trata de espacios privados y aplicaciones comerciales. Si tengo la posibilidad de decidir si mis datos se pueden usar durante 3 días, 3 meses o 3 años, y luego eliminarlos, esto crea libertades que estarían en conflicto con los intereses de las empresas de datos. En cualquier caso, la transparencia de un momento en el que se borran los datos o un “sello de caducidad digital” originan confianza y credibilidad.

Debido a que puede haber conflictos de intereses digitales, se trata, asimismo, del valor de una escalada regulada. No siempre tiene que tratarse del gran umbral de la jurisdicción general. Pero podría haber un defensor del pueblo digital y un tribunal de arbitraje digital que aclaren

los hechos en el sentido de mediación y den valoraciones razonadas. Ciertamente, no dejarían de influir en la práctica digital general y en la jurisprudencia. Además, pueden desarrollarse de forma específica por sector y, en consecuencia, contribuir a la ética profesional digital.

No obstante, puede haber etapas de escalada regulada a nivel de empresas individuales mediante órganos de reflexión adecuados, por ejemplo, en el marco de consejos de ética digital o juntas de ética digital, porque un mundo complejo y emergente también requiere de foros de intercambio regulado sobre diferentes puntos de vista y conflictos de interés que tienen lugar a un nivel más informal que aquel de las autoridades formales.

La combinación de la realización satisfactoria o no satisfactoria de los valores mencionados conduce, en última instancia, a un índice de credibilidad digital. La equidad digital, la credibilidad digital y la creación de confianza van de la mano. En el mejor de los casos, representan el ecosistema emocional de una empresa o institución, y son un indicador de su viabilidad futura.

El objetivo de la equidad digital es, por ende, una tríada de objetivos comerciales comunicativos, éticos y prácticos. Y así como en Alemania existe la fundación pública Stiftung Warentest –para la evaluación objetiva y profesional de ofertas de productos–, podría existir un “Índice de credibilidad digital fundamental” o un “Índice de equidad digital fundamental” para evaluar el comportamiento de las empresas y las organizaciones. Sobre la base de grandes cantidades de datos, una institución de este tipo condensaría las diferentes formas de experiencia del usuario, en términos de equidad digital, y las convertiría en un instrumento para monitorear las ofertas digitales.

La ética digital va más allá de las intuiciones subjetivas. Necesitamos darle la importancia ética al espacio social, contrario a los intereses

puramente comerciales. La ética digital, en consecuencia, se conectará con los valores más importantes para una sociedad humana.

La estrella guía de la humanidad: valores humanos para el mundo digital

La ética digital no cae del cielo, incluye elementos de continuidad y discontinuidad. Este capítulo debería mostrar que podemos ver a la humanidad como un criterio para los desarrollos digitales. En el sentido del *kit* de “artesanía ética” que hemos comentado, hablamos de una reflexión sobre principios con el estándar de una pregunta muy simple: “¿Se trata de promover la humanidad?”.

Hemos discutido en detalle que aquí hay diferentes interpretaciones. Por consiguiente, se requiere de procedimientos para resolver conflictos de valores, a veces también a través de decisiones de mayoría política. Y, sin embargo, a una tecnología, a un programa, a una aplicación, a una instalación, se le puede muy bien preguntar sobre las condiciones en las que nos hace la vida, no solo más fácil y placentera, sino más humana. La pregunta opuesta también se aplica: ¿Bajo qué condiciones el efecto deseado se convierte en el opuesto, de modo tal que nuestra vida finalmente se vuelve inhumana?

La orientación hacia la estrella guía de la humanidad se interpretará de manera distinta según la afiliación sociocultural, profesional, religiosa e ideológica.

No obstante, un criterio común tiene sentido por, al menos, dos razones. Por un lado, sirve como base para evaluar pasos y decisiones específicos, por ejemplo, con miras a la “equidad digital”. Por otro lado, la “humanidad”, en el sentido de una pirámide de valores ordenada jerárquicamente, crea un tirón en el que se puede basar la interpretación de todos los valores posteriores en un sistema de valores. En otras

palabras, el valor central de la humanidad forma el horizonte de interpretación de los demás valores.

Pesar estos valores es, básicamente, un acto de equilibrio, pues las personas consideramos lo que hacemos y cómo lo sopesamos en cada situación de nuestra vida. A menudo se hace recurriendo a las rutinas cotidianas. Pero las actitudes detrás de los hábitos también surgen de acciones concretas sobre las que podemos reflexionar y las que podemos cambiar en caso de duda.

El continuo desde las decisiones cotidianas hasta las grandes cuestiones éticas del mundo, desde la guerra y la paz, desde el principio y el final de la vida, desde el amor y la justicia, puede describirse como una dinámica de acontecimientos en busca del mejor equilibrio posible. A la luz de una interpretación religiosa del mundo, podríamos entender tal dinámica como un intento de estar a la altura de nuestra responsabilidad ante Dios. En el sentido de la ética confuciana, como una búsqueda de equilibrio entre el orden y la armonía celestial y terrenal. Sin embargo, también podemos interpretarla y vivirla sin recurrir a sistemas religiosos específicos.

La cuestión de cómo se pueden equilibrar el orden y la armonía entre las necesidades del individuo y la sociedad es un problema que afecta directamente al mundo digital. Cuando el científico educacional Volker Schubert describe la práctica educativa japonesa como la “puesta en escena de la armonía” (V. Schubert 1992) se trata, además, de la cuestión de la distribución de pesos. ¿Cuánta libertad puede reclamar el individuo? ¿Cuánto control social es razonable?

No es casualidad que la tensión entre libertad y control sea uno de los temas fundamentales de la ética digital. Esta es precisamente la razón por la que los valores de equidad y transparencia son tan relevantes para la acción, porque sirven como filtros para el ejercicio

del control y el poder bajo el lema: el poder ya se ejerce, al menos de forma transparente y según las reglas de la equidad. Esto se aplica tanto a la política como a los actores económicos, es decir, para pequeñas y grandes empresas.

En toda dinámica social, ya sea en una relación de pareja, en la vida laboral o en la gran política, los comportamientos concretos tienen efectos psicológicos. Las consecuencias de la confianza o la destrucción de la confianza son decisivas para la estabilidad de las relaciones, incluyendo la productividad de economías enteras.

La confianza puede describirse como una expectativa positiva del comportamiento de los demás (*cf. G. S. Leventhal, J. Karuza, W. R. Fry 1980, 167-218*). La confianza implica el riesgo de decepcionarse. Por lo tanto, describe una combinación peculiar de expectativas racionales y eventos de relaciones comunicativas en múltiples ciclos de retroalimentación. En sus límites, la confianza se convierte en descuido y atrevimiento, en manía de control y en desconfianza exagerada.

Promover la humanidad significa habilitar la confianza. En última instancia, esta es también la preocupación de la idea de ética global originalmente formulada por Hans Küng (*cf. H. Küng 1990, U. Hemel 2019a*). Me parece fascinante que los representantes de las principales religiones del mundo pudieran ponerse de acuerdo sobre valores comunes, a saber: la veracidad, la justicia, la no violencia, la asociación y la sostenibilidad (*cf. K. J. Kuschel 2019, 39-54; S. Schlensog 2019, 28-38*).

Acá no hay espacio para desarrollar estos valores en detalle. Solo tenemos que demostrar algo: la ética digital puede apoyarse en una rica reserva de valores humanos que, en su totalidad, representan un marco ético para la actividad digital y para cuestiones como los derechos de acceso y participación.

Los conflictos e incertidumbres de valores no significan un obstáculo para la búsqueda de una ética digital adecuada a nuestro tiempo. Efectivamente, los conflictos de valores se pueden resolver mediante el proceso de clarificación y balance de valores, y las incertidumbres se pueden eliminar. La pregunta central que debe plantearse una y otra vez es a dónde queremos llegar como individuo, como empresa, como institución y como sociedad civil global en su conjunto.

Es esta discusión la que debe repetirse para cada aplicación digital individual, pero no carece de objetivo y propósito. Porque con la estrella guía de promover la humanidad mediante la creación de confianza en la equidad y la transparencia, tenemos a mano una brújula válida, relevante y practicable de ética digital que puede proporcionar orientación.

La pregunta es cómo le damos forma a nuestro futuro. Por ello, en el último capítulo de este libro se planteará la cuestión de las utopías y distopías, las esperanzas y las visiones aterradoras del mundo digital.

Referencias bibliográficas

Anderson John R., *The Architecture of Cognition*, Cambridge Mass 1983.

Angel Hans-Ferdinand, *Der religiöse Mensch in Katastrophenzeiten, Religionspädagogische Perspektiven kollektiver Elendsphänomene*, Frankfurt/M.: Lang 1996 (Habilitation 1994, Universität Regensburg).

Anzenbacher Arno, *Einführung in die Ethik*, 3. Auflage, Düsseldorf: Patmos 2003.

Asimov Isaac, *Meine Freunde, die Roboter*. München: Heyne 1982.

Bedford-Strohm Jonas, Filipovic Alexander, *Mediengesellschaft im Wandel, Theorien, Themen, ethische Herausforderungen*, in: Gotlind Ulshöfer, Monika Wilhelm (Hrsg.), *Theologische Medienethik im digitalen Zeitalter*, Stuttgart: Kohlhammer 2020, S. 47-63.

Bielefeldt Heiner, *Philosophie der Menschenrechte, Grundlagen eines weltweiten Freiheitsethos*, Darmstadt: Wissenschaftliche Buchgesellschaft 1998.

Binder Franziska, Werner Amelie, Fülöp Rea, Parzer Peter, Koenig Julian, Resch Franz, Kaess Michael, *Indizierte Stressprävention bei Jugendlichen im Gruppensetting-Manual für einen Akzeptanz- und Commitmenttherapie-basierten Ansatz*, in: *Praxis der Kinderpsychologie und Kinderpsychiatrie*, Bd. 69, 2020, S. 183-202.

- Bohlken Eike, Thies Christian (Hrsg.),** Handbuch Anthropologie, Der Mensch zwischen Natur, Kultur und Technik, Stuttgart: J. B. Metzler 2009.
- Bollnow Otto F.,** Wesen und Wandel der Tugenden, Frankfurt/M.: Ullstein 1958.
- Bostrom Nick,** Superintelligenz, Szenarien einer kommenden Revolution, Berlin: Suhrkamp 2016.
- Braun Ulrike,** Exzessive Internetnutzung Jugendlicher im familialen Kontext, Analysen zu Sozialschicht, Familienklima und elterlichem Erwerbsstatus, Wiesbaden: Springer Fachmedien 2014.
- Burns James H.,** Happiness and Utility, Jeremy Bentham's Equation, in: Utilitas 17, 2005, Nr. 1, S. 46-61.
- Copray Norbert,** Fairness, Der Schlüssel zu Kooperation und Vertrauen, Gütersloh: Gütersloher Verlagshaus 2010.
- Crane Tim,** Die Bedeutung des Glaubens, Religion aus der Sicht eines Atheisten, 2. Aufl. Berlin: Suhrkamp 2019.
- Dechesne Francien, Warnier Martijn, Van den Hoven Jeroen,** Ethical Requirements for Reconfigurable Sensor Technology: A Challenge for Value Sensitive Design, in: Ethics and Information Technology 15, 2013, no. 3, S. 173-181.
- Demmer Klaus,** Deuten und Handeln, Grundlagen und Grundfragen der Fundamentalmoral, Freiburg/Breisgau: Herder 1985.
- De Waal Frans,** Der Mensch, der Bonobo und die zehn Gebote, Moral ist älter als Religion, Stuttgart: Klett-Cotta 2015.

- Dickel Sascha**, Enhancement-Utopien, Soziologische Analysen zur Konstruktion des Neuen Menschen, Baden-Baden: Nomos 2011 (Dissertation Universität Bielefeld 2010).
- Edelstein Wolfgang, Keller Monika (Hrsg.)**, Perspektivität und Interpretation, Beiträge zur Entwicklung des sozialen Verstehens, Frankfurt/M: Suhrkamp 1982.
- Eggensperger Thomas, Engel Ulrich**, Bartolomé de las Casas, Dominikaner-Bischof, Verteidiger der Indios, 2. Auflage, Mainz: Matthias-Grünewald 1992.
- Eysenck Michael W., Keane Mark T.**, Cognitive Psychology, A Student's Handbook, 6. Auflage, Hove-New York: Psychology Press 2010.
- Fink Franziska, Moeller Michael**, Purpose Driven Organisations, Sinn, Selbstorganisation, Agilität, Stuttgart: Schaefer-Pöschel 2018.
- Friedman Batya, Hendry David G.**, Value Sensitive Design, Shaping Technology with Moral Imagination, Cambridge/Mass.: MIT Press 2019.
- Gadamer Hans-Georg**, Wahrheit und Methode, 7. Auflage, Tübingen: Mohr Siebeck 1960.
- Gaschler Robert, Frensch Peter A.**, Kognitivismus, in: Eike Bohlken, Christian Thies (Hrsg.), Handbuch Anthropologie, Der Mensch zwischen Natur, Kultur und Technik, Stuttgart: Metzler 2009, S. 157-163.
- Gehlen Arnold**, Der Mensch, Seine Natur und seine Stellung in der Welt, (ursprünglich 1940), hrsg. von Karl-Siegbert Rehberg, Frankfurt/M.: Vittorio Klostermann 1983.

- Grimm Petra, Keber Tobias O., Zöllner Oliver (Hrsg.)**, Digitale Ethik, Leben in vernetzten Welten, Stuttgart: Reclam 2019.
- Hagencord Rainer**, Die Würde der Tiere, Eine religiöse Wertschätzung, Gütersloh: Gütersloher Verlagshaus 2011.
- Hayes Steven C. (Hrsg.)**, Relational Frame Theory, A Post-Skinnerian Account of Human Language and Cognition, New York: Kluwer 2001.
- Hayes Stephen (sic!) C., Strosahl Kirk D., Wilson Kelly G.**, Akzeptanz- & Commitment Therapie, Achtsamkeitsbasierte Veränderungen in Theorie und Praxis, Paderborn: Junfermann 2014.
- Hemel Ulrich**, Ziele religiöser Erziehung, Beiträge zu einer integrativen Theorie, Frankfurt/M.: Peter Lang 1988 (Habilitation Universität Regensburg 1988).
- Hemel Ulrich**, Wert und Werte, Ethik für Manager, Ein Leitfaden für die Praxis, 2. Auflage, München: Hanser 2007 (*Valor y valores: Ética para directivos*, Barcelona 2007).
- Hemel Ulrich, Fritzsche Andreas, Manemann Jürgen (Hrsg.)**, Habituelle Unternehmensethik, Von der Ethik zum Ethos, Baden-Baden: Nomos 2012.
- Hemel Ulrich**, Die Wirtschaft ist für den Menschen da, Vom Sinn und der Seele des Kapitals Ostfildern: Patmos 2013 (*Economía para el ser humano, Sentido y alma del capital*. Siglo del Hombre, Bogotá 2016).
- Hemel Ulrich**, Tolerance and the “right of error” in a world of religious plurality, in: Markus Krienke, Elmar Kuhn (Hrsg.), Two indispensable topoi of interreligious dialogue. New “languages” far beyond the dead ends of dialogue, Wien 2017, S. 65-78.

- Hemel Ulrich**, Heimat und personale Selbstbildung. Eine pädagogische Reflexion, in: Ulrich Hemel, Jürgen Manemann (Hrsg.), Heimat finden-Heimat erfinden. Politisch-philosophische Reflexionen, Paderborn: Wilhelm Fink 2017, S. 157-173.
- Hemel Ulrich**, Was heißt eigentlich digitale Fairness? <https://www.institut-fuer-sozialstrategie.de/2016/11/10/was-heisst-eigentlich-digitale-fairness/> Veröffentlicht am 10. November 2016; abgerufen am 16. April 2020 um 9.23 h.
- Hemel Ulrich (Hrsg.)**, Weltethos für das 21. Jahrhundert, Freiburg: Herder 2019 (U. Hemel 2019a).
- Hemel Ulrich**, Mentale Architektur und Wirtschaftsanthropologie - eine Zukunftsaufgabe, in: Sebastian Kiessig, Marco Kühnlein (Hrsg.), Anthropologie und Spiritualität für das 21. Jahrhundert, Festschrift für Erwin Möde, Regensburg: Pustet 2019, S. 335-350.
- Hemel Ulrich**, Was ist Wahrheit? Die Wahrheitsfähigkeit der Welt und die Weltethos-Idee, 2020, in: www.institut-fuer-sozialstrategie.org, abgerufen am 25.4.2020.
- Joas Hans**, Die Entstehung der Werte, Frankfurt/M.: Suhrkamp 1997.
- Joisten Karen**, Philosophische Hermeneutik, Berlin: De Gruyter, 2009.
- Kant Immanuel**, Grundlegung zur Metaphysik der Sitten, in; Kants Werke, Akademie-Textausgabe (AA), Bd. IV, Berlin 1968.
- Kirchschläger Peter**, Digital Transformation and Ethics, Ethical Considerations on the Robotization and Automatization of Society and Economy and the Use of Artificial Intelligence, Baden-Baden: Nomos 2020.

- Küng Hans**, Projekt Weltethos, München: Piper 1990.
- Kuschel Karl-Josef**, Weltreligionen im Dialog über Weltprobleme, Die Erklärung des Parlaments der Weltreligionen in Chicago, in: U. Hemel (Hrsg.), Weltethos für das 21. Jahrhundert, Freiburg/Br. 2019, 39-54.
- Leventhal G. S., Karuza Jurgis, Fry W. R.**, Beyond Fairness, A Theory of Allocation Preferences, in: Gerold Mikula (Hrsg.), Justice and Social Interaction, New York: Springer 1980, S. 167-218.
- Lorenz Konrad**, Vergleichende Verhaltensforschung. Grundlagen der Ethologie, Wien/New York: Springer 1978.
- MacIntyre Alasdair**, Der Verlust der Tugend, Zur moralischen Krise der Gegenwart, Frankfurt: Suhrkamp 1995 (engl. 1981).
- Margalit Avishai**, Politik der Würde, Über Achtung und Verachtung, 2. Auflage, Berlin: Suhrkamp 1997.
- Misselhorn Catrin**, Grundfragen der Maschinenethik, Stuttgart: Reclam 2018.
- Mourkogiannis Nikos**, Purpose, The Starting Point of Great Companies, New York/Basingstoke: Palgrave Macmillan 2006.
- Nida-Rümelin Julian, Weidenfeld Nathalie**, Digitaler Humanismus, Eine Ethik für das Zeitalter der Künstlichen Intelligenz, München: Piper 2018.
- Nissenbaum Helen**, Computing and Accountability, Communications of the ACM, 37, Nr.1, 1994, S. 73-80.

- Parlament der Weltreligionen**, <https://parliamentofreligions.org/parliament/global-ethic/fifth-directive> und <https://parliamentofreligions.org/publications/towards-global-ethic-initial-declaration>, abgerufen am 11.4.2020, 19.34 h.
- Pieper Josef**, Gesammelte Werke, hrsg. von Berthold Wald, Band 4: Schriften zur Philosophischen Anthropologie und Ethik: Das Menschenbild der Tugendlehre, Hamburg: Felix Meiner 1996.
- Pirner Manfred L., Lähnemann Johannes, Bielefeldt Heiner (Hrsg.)**, Human Rights and Religion in Educational Contexts, Cham: Springer International Publishing 2016.
- Roth Erwin, Holling Heinz**, Sozialwissenschaftliche Methoden, Lehr- und Handbuch für Forschung und Praxis, 5. Auflage, Berlin: De Gruyter 2018.
- Scheler Max**, Der Formalismus in der Ethik und die materiale Wertethik, hrsg. von Christian Bermes, Hamburg 2014 (Original: 1913-1916).
- Schlensof Stephan**, Das Weltethos-Projekt, Drei Jahrzehnte globaler gesellschaftlicher Dialog, in: Ulrich Hemel (Hrsg.), Weltethos für das 21. Jahrhundert, Freiburg/Br.2019, S. 28-38.
- Schubert Volker**, Die Inszenierung der Harmonie, Erziehung und Gesellschaft in Japan, Darmstadt: Wissenschaftliche Buchgesellschaft 1992.
- Schütz Alfred**, Der sinnhafte Aufbau der sozialen Welt, Eine Einleitung in die verstehende Soziologie, Frankfurt am Main: Suhrkamp 1974 (ursprünglich Wien: Springer 1932).

- Schütz Alfred**, Theorie der Lebensformen, hrsg. von Ilja Srubar, Frankfurt/M.: Suhrkamp 1981.
- Singer Peter**, Animal Liberation, Die Befreiung der Tiere, Eine neue Ethik zur Behandlung der Tiere, München: Fischer 1982 (2. Aufl. 2015).
- Spiekermann Sarah**, Digitale Ethik, Ein Wertesystem für das 21. Jahrhundert, München: Droemer 2019.
- Ströker Elisabeth (Hrsg.)**, Lebenswelt und Wissenschaft in der Philosophie Edmund Husserls. Frankfurt am Main: Klostermann 1979.
- Taleb Nassim Nicholas**, Der Schwarze Schwan, Die Macht höchst unwahrscheinlicher Ereignisse, München: Hanser 2008.
- Tiedemann Paul**, Was ist Menschenwürde? Eine Einführung, Darmstadt: Wissenschaftliche Buchgesellschaft 2006.
- Timmermans J., Zhao Y., Vanden Hoven J.**, Ethics and Nanopharmacy: Value Sensitive Design of News Drugs, in: NanoEthics 5, 2011, S. 269-283.
- Ulshöfer Gotlind, Wilhelm Monika (Hrsg.)**, Theologische Medienethik im digitalen Zeitalter, Stuttgart: Kohlhammer 2020.
- Wagner Hermann**, Der Mensch als Tier, die biologische Perspektive, in: Ulrich Lüke, Georg Souvignier (Hrsg.), Der Mensch - ein Tier. Und sonst? Interdisziplinäre Annäherungen, Freiburg/Br.: Herder 2020, S. 45-65.
- Ziebertz Hans-Georg (Hrsg.)**, Menschenrechte, Christentum und Islam, Berlin: LIT 2010.



2





Religión digital y humanidad digital

Las cuestiones religiosas producen repetidamente cierto malestar en la vida social y, en especial, en los contextos académicos, entre los científicos naturales, sociales y económicos. No se puede negar que existe la religión, pero cualquier debate sobre ella puede terminar fácilmente en una zona tabú.

A esto contribuye la preocupación por entrar en el área de las sensibilidades emocionales. Después de todo, a menudo no se sabe qué opinión tiene el interlocutor y con qué intensidad se encuentra ideológica y religiosamente ligado a una creencia. La otra cara de la moneda es la suposición científica de que la religión y las concepciones religiosas difícilmente pueden discutirse de forma científica.

El filósofo británico Tim Crane se expresó de manera muy reflexiva desde una perspectiva atea (T. Crane 2019). El impulso antirreligioso de las fuerzas científicas reconoce que las ciencias empíricas –como la sociología de la religión, la psicología religiosa y la historia de la religión y de la teología– deben tomarse en serio de acuerdo con los estándares científicos. Las propias concepciones religiosas, sin embargo, son relegadas apresuradamente al reino de la opinión subjetiva, quizás incluso al esoterismo, la superstición y la pura especulación.

Las reservas correspondientes también se aplican en el ámbito de la religión digital. En primer lugar, la transformación digital no necesita una superestructura religiosa o ideológica. Por lo tanto, para muchas personas, tratar las cuestiones de la religión digital con todas sus formas extrañas y bizarras no es importante o, en el mejor de los casos, irrelevante.

Por otro lado, la velocidad y la calidad de la transformación digital plantean nuevas interrogantes, como la demarcación entre personas y máquinas. Además, el mundo de las computadoras, los cibernéticos y la inteligencia artificial estimula la imaginación de novelistas, cineastas y productores de videojuegos. De esta actividad surge lo que contiene el término “ciencia ficción”: una mezcla de ciencia y ficción, de tecnología y fantasía. Parte de dicha fantasía se centra, a su vez, en la posible superioridad de las máquinas superinteligentes o “seres máquina”.

La conexión entre la idea de seres superiores y máquinas digitales sobre la base de la inteligencia artificial lleva, entonces, al área de la especulación filosófica y teológica y a líneas de pensamiento que conducen, en cierta medida, al área de la religión y la religiosidad.

En 1986, el autor estadounidense John Updike (1932-2009) escribió una novela llamada *La versión de Roger*, en la que, entre otras cosas, se alude a una prueba de Dios basada en el procesamiento electrónico de datos (cf. J. Updike 1990). Alrededor de 30 años después, la novela *Máquinas como yo*, del británico Ian McEwan, aborda una cuestión completamente diferente: ¿Puede el androide Adam pensar, sentir y actuar como un humano, para que no deba ser destruido, sino que se le otorguen los derechos de una persona? Por consiguiente, la destrucción de la máquina sería tan problemática como un asesinato (I. McEwan 2019).

La imagen de Dios: la dignidad de las personas y las máquinas

La cuestión de la personalidad de un androide controlado por inteligencia artificial no es solo un asunto de novelas. Más bien, nos encamina rápidamente hacia cuestiones legales, filosóficas y religiosas.

Si vemos la dignidad humana como el punto de partida para los sistemas de valores éticos y una moralidad cotidiana práctica para las personas en la sociedad civil global y local en el siglo XXI, surge la cuestión de la justificación de la dignidad humana.

En el contexto religioso, las interpretaciones cristiana, judía e islámica tratan del hombre a imagen de Dios. El texto clásico en Génesis, Capítulo 1, versículo 26, “Entonces dijo Dios: ¡Hagamos a las personas a nuestra imagen como a nuestra semejanza!”, es la base de las ideas sobre la justificación de la dignidad humana a la imagen de Dios. Incluso la palabra “educación” se relaciona con la “formación” del ser humano en la dirección de esta “imagen de Dios”, como muestra el historiador de la educación religiosa Hans Schilling en su aún clásico estudio de la historia del concepto de “formación” (H. Schilling 1961).

Por ende, el ser humano tiene una dignidad fundamental e inalienable, que se expresa en su personalidad. Si una persona es una persona, entonces con Immanuel Kant es un fin en sí mismo: no debe convertirse ella misma en un fin. Si una persona es una persona, es capaz de controlarse y asumir la responsabilidad. Por lo tanto, puede ser considerada responsable, pero también rendir cuentas en el sentido de presentar buenas razones para sus acciones.

Nuestro sistema legal democráticamente legitimado descansa sobre esta base. Los motivos religiosos pasan a un segundo plano porque el discurso de los derechos humanos y la apelación a la dignidad humana ya se han desarrollado, durante mucho tiempo, independientes de los argumentos teológicos y las líneas de razonamiento religiosas.

La personalidad y, por consiguiente, la responsabilidad personal siguen siendo una base indispensable para dar forma a la sociedad civil mucho más allá de Europa. A menudo, las raíces religiosas de la personalidad y la dignidad humana a imagen de Dios ya no se tienen en cuenta hoy. No obstante, poseen una fuerza definitoria que también afecta al campo de la religión digital.

En el texto bíblico, la creación del hombre por Dios va de la mano con la llamada “comisión de creación”, que está formulada de la siguiente manera: “Y los bendijo Dios, y les dijo: Fructificad y multiplicaos; llenad la tierra y sojuzgadla” (Gen 1, 28).

En tiempos de crisis climática y declive de la biodiversidad, la gran mayoría de los teólogos interpretan el texto de modo más cauteloso y crítico que en siglos anteriores. Se refieren, por ejemplo, a la responsabilidad de la preservación de la creación –que es parte de la comisión de creación–, para así lograr la conectividad de su interpretación bíblica con el movimiento ecológico actual.

Sin embargo, la historia bíblica sobre la creación del hombre a imagen de Dios plantea preguntas duraderas, ya que la personalidad humana no se puede tener sin libertad. La libertad, empero, implica el poder de actuar, pero también la responsabilidad. La dignidad humana y la imagen de Dios marchan a la par. La creación del hombre como persona en libertad y responsabilidad va de la mano con las consecuencias, a veces dramáticas, de la libertad en la *Conditio humana*, con la culpa y la maldad, con la impotencia y la desesperación, con el infinito y los duros límites de la finitud.

La cuestión metafísica de la transformación digital se puede deducir fácilmente de estas consideraciones: ¿Es la *Conditio humana* también una *Conditio machinae*? Es decir, ¿las máquinas inteligentes



Las raíces religiosas de la personalidad y la dignidad humana a imagen de Dios poseen una fuerza definitoria que también afecta al campo de la religión digital.

y las entidades digitales tienen el atributo de la persona, con todas sus consecuencias?

Entonces, no serían los programadores de inteligencia artificial sino los seres máquina basados en esta IA los responsables, pero además serían culpables, poderosos y, al mismo tiempo, impotentes como nosotros, los humanos. En otras palabras, serían personas creadas a imagen del hombre. Y de la personalidad de las máquinas habría que inferir su dignidad personal. Por ende, la destrucción de una máquina sería asesinato u homicidio involuntario, al igual que, en la actualidad, el asesinato de una persona se llama “asesinato” u “homicidio involuntario”.

Estas líneas de pensamiento son fascinantes porque, así como debemos aprender a complementar la diferencia entre humanos y animales con la distinción entre humanos y máquinas, de cierta forma duplicaríamos la historia de la creación. Dios crea al hombre a su propia imagen. Por otra parte, el ser humano crea seres digitales según su propia imagen.

Varias fantasías van de la mano con una nueva historia de creación. La doctrina clásica de las propiedades de Dios con la atribución de omnipotencia y omnisciencia está siendo redescubierta en el campo

digital. Si uno sigue las teorías de la conspiración cotidiana, escucha una y otra vez que los grandes actores digitales “saben todo sobre nosotros” y “nos controlan”.

Si observa más de cerca la realidad, es fácil ver que no son las máquinas digitales las que conocen y controlan, sino las personas, las organizaciones y los poderes políticos quienes saben cómo utilizar dichas máquinas. El miedo a un control cada vez mayor no es sin causa y razón: al fin y al cabo, las nuevas posibilidades tecnológicas no son éticamente neutrales. Por el contrario, pueden llevar a la tentación y al hecho de una centralización y omnipresencia aún mayor de la vigilancia digital.

Pero ni la religión digital ni la destrucción de las computadoras pueden ayudar. Más bien, ¡las preguntas sobre la calidad de la política y el objetivo de la coexistencia social en una democracia conformada digitalmente deben formularse de nuevo! Como resultado, se buscan nuevos modelos de autodeterminación y libertad en el mundo digital, un aspecto que es de particular importancia.

Una máquina, por otro lado, difícilmente se preocuparía por su propia monitorización digital. Tampoco es probable que tenga sentido hablar de los estados emocionales reales de las máquinas de un modo que señale la indiferencia como un sentimiento. No obstante, dado que las máquinas están diseñadas y son usadas para fines específicos, no tiene sentido otorgarles derechos personales y la dignidad de una persona. ¡Las máquinas no poseen dignidad humana ni una dignidad especial propia!

Las máquinas pueden causar destrucción o facilitar nuestra vida diaria. Pueden romperse o ser duraderas. Pueden ayudar a cambiar nuestro comportamiento, y se pueden utilizar para todo tipo de propósitos, claros y oscuros. Todas las máquinas, sin embargo, son

objeto de producción técnica con participación humana, directa o indirecta. Por lo tanto, las máquinas no son culpables. Pueden salirse de control, pero como máquinas no se les responsabiliza: no es la máquina la que va a la cárcel, sino su diseñador o su cliente.

Cuando el vuelo 302 de Ethiopian Airlines de Addis Abeba a Nairobi se estrelló poco después del despegue, el 10 de marzo del 2019, y murieron 157 personas, comenzó una intensa solución de problemas. Resultó que la programación basada en inteligencia artificial del Boeing 737 MAX 8 no permitía la intervención humana en suficiente medida. Entonces, se detuvo la entrega de este tipo de aviones, con enormes consecuencias económicas para el fabricante. Hasta el momento no ha habido fallos judiciales al respecto, pero está claro que nadie tenía la idea de responsabilizar al programa defectuoso o “castigarlo”.

Por consiguiente, cualquier forma de ética de la máquina solo puede utilizar el concepto de responsabilidad ética como analogía, como transferencia pictórica. Las máquinas digitales calculan. El resultado de su cálculo puede tomar la forma de una instrucción con la consecuencia de una acción real. Los resultados de los cálculos de inteligencia artificial pueden simular decisiones éticas. La calidad de esta simulación puede ser incluso mejor que la de los tomadores de decisiones humanos.

No obstante, la responsabilidad sigue siendo del ser humano, a través del acto de aceptación, implícita o explícita, del resultado de la máquina en el propio contexto de acción. El mundo digital no puede ni debe liberarnos de nuestra responsabilidad ética: porque las personas son personas; las máquinas no lo son.

Superinteligencia digital e inmortalidad digital

A primera vista, puede parecer tranquilizador entender los límites de la fantasía y de la imaginación humana alrededor de sus creaciones. Efectivamente, no asignamos a las máquinas digitales la libertad de tomar decisiones o de actuar. Por ende, nuestras creaciones digitales no tienen ni la libertad ni la responsabilidad de las mismas personas humanas.

No es una coincidencia que el exitoso libro de Yuval Harari lleve el título *Homo Deus*, que significa “el hombre como Dios” o “el hombre parecido a un dios” (Y. Harari 2017). Se trata del futuro de las personas que se reinventan con la ayuda de las nuevas tecnologías y de máquinas que hemos creado nosotros y que pueden hacer muchas cosas mejor que los seres humanos. La búsqueda de la felicidad, la salud y el poder puede ayudar a mejorar nuestras vidas, pero también sumirnos en la infelicidad.

Por supuesto, esta intuición nos ha acompañado desde el mito de Prometeo, quien robó el fuego de los dioses para los humanos, pues los aspectos de la tecnología que mejoran la vida y que son destructivos suelen estar muy próximos.

La tecnología digital, en particular, sirve a la expectativa de un mayor rendimiento cognitivo. Se convierte en la expresión de las utopías de mejora, como dice Sascha Dickel (S. Dickel 2011). La lógica de mejorar el rendimiento físico y mental de las personas siempre ha formado parte de su historia, desde el bastón blanco hasta la ayuda visual, desde el audífono hasta el marcapasos.

La misma lógica de aumento real o supuesto del rendimiento incluye medicamentos como Ritalin o Viagra, cirugía estética y sangre autóloga para dopaje en el deporte, pero también sustancias psicotrópicas, como el LSD y otras drogas, aunque las aplicaciones antes citadas suelen ser más críticas y ambivalentes en la sociedad.

El “dopaje cerebral” a través de chips cerebrales, es decir, implantes que funcionan digitalmente, ya no está fuera de nuestro alcance. Los marcapasos cerebrales en la enfermedad de Parkinson son un ejemplo médico.

Debido a que la gente ha utilizado la tecnología desde el principio, la antropología siempre puede percibirse como antropología técnica. Las transiciones en el uso de baja tecnología, como un bastón, a alta tecnología, como un neuroestimulador o un marcapasos cerebral, son fluidas. Una cultura de la digitalidad (F. Stalder 2016) presupone, en todo caso, el correspondiente desarrollo técnico, incluyendo sus peligros (cf. J. Lanier 2014).

Cualquiera que interprete el término “ciborg” como un híbrido humano-máquina podrá considerar a todas las personas que dependen de prótesis, implantes cocleares y otras ayudas tecnológicas médicas, como ciborgs. En los casos más raros, esta idea corresponde a la imagen que aquellas tienen de sí mismas. Sin embargo, el límite entre la compensación de las debilidades físicas y las medidas para aumentar el rendimiento físico y mental ya no puede definirse claramente.

En el contexto de una “crítica de la razón digital” surge la gran pregunta sobre en qué punto termina la idea de incrementar el desempeño y cuándo las personas, tal como las conocemos, se convierten en artefactos o, incluso, en “objetos de diseño” (cf. B. Göcke, F. Meier-Hamidi 2018). Entonces, estamos hablando de la era posthumana o del transhumanismo (cf. B. Flessner, D. Spreen et al. 2018).

El *biobacking* es una forma especial de cruce de fronteras (cf. H. Charisius, R. Friebe, S. Karberg 2013). Es la conexión de funciones fisiológicas y biológicas del cuerpo humano con interfaces digitales, por ejemplo, a través de chips, sensores o imanes implantados. La curiosidad experimental suele estar en segundo plano; a veces se trata

de los modos de autooptimización del propio cuerpo ya comentadas, también mediante experimentos con secuencias de ADN. El límite entre la experimentación lúdica y la fijación ideológica, por un lado, y los intereses comerciales o transhumanistas-ideológicos, por el otro, parece fluido.

La combinación de neurociencias y tecnología digital es nueva. Ya hoy en día se busca una mejora en la comunicación humana utilizando una interfaz cerebro-computadora en pacientes paralizados, entrenando las funciones de movilidad a través de la actividad eléctrica del cerebro (cf. *K. H. Pantke 2010*). No obstante, esta es tecnología médica sin connotaciones religiosas.

Sin embargo, la sola idea de una interfaz cerebro-máquina estimula mucho la imaginación humana. Numerosas películas y videojuegos sobre cibernéticos –controlados desde el exterior– dan testimonio de ello. Se trata de poder y manipulación, pero también de megalomanía y, específicamente, de ideas religiosas. Si no hay límite para el aumento del rendimiento cognitivo, terminamos en el campo del infinito, incluso en la inmortalidad digital.

Una faceta especial al considerar los rasgos especulativos y parcialmente religiosos de la tecnología digital es la creencia en la singularidad o superinteligencia, que será muy superior a los humanos. Los protagonistas de esta idea son, entre otros, el filósofo sueco Nick Bostrom (*N. Bostrom 2005, 11-30*) y el empresario informático estadounidense Ray Kurzweil (2005), que fundó su propia Singularity University en 2008, en Silicon Valley.

El término “singularidad” se utiliza para describir el momento en el que emerge una superinteligencia posthumana que puede dominar a las personas y remodelar totalmente nuestra civilización (cf. *S. Dickel 2011, 285*). Si uno piensa en el tema con mayor intensidad, aparece

el punto de vista de la preocupación o el miedo a la superioridad de las máquinas sobre las personas. Aquí, la lógica del aumento de las capacidades cognitivas respaldadas digitalmente se combina con la cuestión del poder o, incluso, del control de la humanidad.

Como suele ocurrir con la conexión entre utopía y apocalipsis, con posibilidades inimaginables y riesgos existenciales, también hay esperanza de cura. Una de ellos es la carga digital de todo el cerebro humano y su completa digitalización, para que podamos lograr una especie de inmortalidad digital (*cf. O. Krüger 2004*).

Para la mayoría de las personas, no obstante, una existencia puramente digital o “cerebral” no corresponde con sus ideas de la buena vida. Pero también se ofrece un remedio. Una posibilidad con un toque parareligioso es congelar el cuerpo con la idea de controlar el mundo físico sobre una base molecular. La Alcor Life Extension Foundation y el Cryonics Institute ya han criopreservado a varios cientos de personas después de su muerte (*cf. S. Dickel 2011, 254*). Los contratistas criónicos esperan que algún día sus cuerpos puedan ser resucitados. Es, literalmente, una esperanza de resurrección física.

Tales formas de autotranscendencia infinita están vinculadas con una semántica que recuerda el lenguaje religioso. Por ejemplo, se trata de la liberación de la naturaleza, de la fantasía de omnipotencia de la evolución controlada por el ser humano y de la conexión entre la utopía técnica y social. Así como se puede descifrar el genoma humano, se puede descifrar el cognomen humano, o sea, la totalidad del contexto estructural del cerebro.

En mi propia vida puedo recordar muy bien el clima revolucionario en algunos círculos estudiantiles hacia fines de la década de 1970. Para comprender mejor los contextos utópicos, en mi tesis de calificación de licenciatura había estudiado intensamente a Herbert

Una faceta especial al considerar los rasgos especulativos y parcialmente religiosos de la tecnología digital es la creencia en la singularidad o superinteligencia.



Marcuse, uno de los filósofos más conocidos de la Escuela Crítica de Frankfurt (*cf. H. Marcuse 1967*). Pude mostrar cuán fuertemente sus ideas revolucionarias sociales fueron influenciadas por las ideas religiosas del mesianismo judío.

Las esperanzas e ilusiones social-revolucionarias de aquella época han dado paso hoy a una cierta desilusión en amplios sectores de la sociedad. Mientras tanto, la conexión de una tradición política individualista con utopías científica y técnicamente justificables parece ser compatible en círculos técnicamente eufóricos. Entonces, vivimos una especie de utopía liberal.

Incluso el objeto de la esperanza aparentemente religiosa se convierte en objeto de una mejora interior: no solo la sociedad, sino toda la naturaleza humana, se transforma en el objetivo de la utopía. El desarrollo digital es solo una parte de esto, pero con miras a cargar todo el cerebro (*uploading*) o a mejorar el rendimiento cerebral a través de chips implantables, ciertamente es una parte esencial de dichas esperanzas.

Partiendo del supuesto de que una persona procesará 10 terabytes de información a lo largo de su vida, es posible imaginar que, en el 2040, tal volumen de datos cabrá en un microchip. Esto fue informado en el

2003 por el autor Christoph Keller sobre una conferencia dada en 1996 por Chris Winter, entonces jefe del departamento de Vida Artificial en el laboratorio de investigación de British Telecom. Con esos microchips, por ejemplo, a una persona con la enfermedad de Alzheimer se le podría implantar otro chip cerebral (*cf. Ch. Keller 2003, 39*).

La contextualización económica, política y social de aquellas utopías tecnológicas se ignora a menudo en el contexto de un individualismo y utilitarismo fuertemente anglosajones. Al mismo tiempo, la línea entre la esperanza técnica para el futuro y la expectativa de redención con un toque de religión se está volviendo borrosa. El problema no es solo la creencia, a veces acrítica, en lo que es técnicamente factible, sino más bien eludir la cuestión del control social, las consecuencias ecológicas y la cultura de la toma de decisiones políticas. Para decirlo en pocas palabras: cualquiera que quiera congelarse tras su muerte y tener su cerebro cargado de antemano debe poder permitírselo económicamente.

Esta exclusividad de élite de la tecnología social digital contrasta con las expectativas de salvación universal de las grandes religiones del mundo. Según el estado actual de las cosas, el infinito apoyado digitalmente requiere de mecanismos de asignación y selección social que al menos puedan ser cuestionados, porque representan una utopía de exclusión social, no de inclusión. Quien procede de esa manera en la esfera religiosa suele elegir la forma social de la secta, no la de la iglesia cristiana o la comunidad religiosa integral, como en la umma islámica.

En contraste con tal individualismo de salvación digital, la fenomenología del mundo digital muestra características de una religión cotidiana universal. Esta observación se profundizará en la siguiente sección.

La religión digital como esperanza de salvación digital

En su libro que se titula *Google Our* (Google Nuestro), Christian Hoffmeister describe la cultura de Silicon Valley utilizando el ejemplo del gigante tecnológico Google (Ch. Hoffmeister 2019). La empresa Google –pero también Facebook y Apple– se compara con una iglesia cuyos seguidores son los usuarios.

Su creencia está determinada por la convicción de que todos los problemas del mundo pueden resolverse con datos, algoritmos y supercomputadoras adecuados, pues la creencia digital incluye la convicción religiosa de que el mundo es fundamental y completamente calculable y predecible.

En una parodia al padrenuestro cristiano, la oración del creyente digital dice lo siguiente (*ibid.*, 2):

“Google Nuestro
Que estás en la Nube,
Santificada sea tu búsqueda,
Venga a nosotros tu buscador,
Hágase tu algoritmo,
Así en las laptops como en los celulares.

Danos hoy nuestros *likes* de cada día,
Y perdona nuestros *dislikes*,
Así como nosotros perdonamos a nuestros *haters*.
No nos dejes caer en las páginas nada relevantes,
Y líbranos de nuestro propio conocimiento.
Porque tuya es la red y la omnisciencia y la singularidad en la eternidad.
Amén”.

La expectativa digital de salvación se refleja en esta “oración” porque el propio pensamiento se valora menos que las ofertas digitales y la creencia en el progreso tecnológico se capitaliza.

Anthony Levandowski, quien fue contratado por Google como desarrollador de Google Street View, va aún más lejos. Fundó la iglesia informática Way of the Future y la registró en los EE. UU. en agosto del 2017. Un informe al respecto se publicó en la revista Techbook el 17 de noviembre del 2017 (<https://www.techbook.de/easylife/web/religion-kuenstliche-intelligenz-way-of-the-future>, consultado el 2 de mayo de 2020, a las 8:33 p.m.).

Para este “fundador de una religión”, la “realización, aceptación y adoración de una deidad basada en la inteligencia artificial” está en primer plano. Es importante construir una relación positiva con los sistemas todopoderosos que controlarán el mundo en el futuro (*ibid.*).

No es difícil diferenciarse de las polémicas pseudorreligiosas contra las corporaciones digitales, así como de las extrañas ideas sobre la divinidad de la inteligencia artificial. Sin embargo, es parte de nuestro mundo que el rápido cambio –especialmente en el campo de la inteligencia artificial– puede conducir a temores apocalípticos de sentirse abrumado y a expectativas religiosas de salvación. Los atributos clásicos de Dios, empero, se reducen mucho a los aspectos de omnisciencia y omnipotencia. El amor y la misericordia, que en el cristianismo, el judaísmo y el islam son características esenciales de la relación de Dios con las personas humanas, no juegan un papel.

Este estrechamiento cognitivista de la religión recuerda algunas palabras de la gnosis. Se trata de la autorredención de los seres humanos mediante un conocimiento superior, es decir, una “competencia” con el mundo del pensamiento de las religiones reveladas, que esperan la redención y la salvación a través de la intervención de Dios desde el

exterior (cf. *W. Kosack 2014*). Las concepciones gnósticas, esotéricas y, en el sentido más amplio, parareligiosas se combinan luego con las ideas digitales de salvación.

No obstante, hasta ahora, las iglesias digitales, los cultos digitales y las deidades no han tenido mayor éxito. Por el contrario, el uso del mundo digital ha penetrado desde hace mucho tiempo en la práctica de las comunidades religiosas e iglesias establecidas. Por ende, la práctica religiosa en el mundo digital y las teologías explícitamente digitales merecen una atención especial.

Práctica religiosa en mundos digitales y teología digital

En su libro *Religión digital*, la científica estadounidense de la comunicación Heidi Campbell se refiere al uso de la tecnología digital por parte de las comunidades religiosas (H. Campbell 2012). Sobre todo a raíz de la pandemia del coronavirus del 2020, el uso de formas digitales de adoración se ha convertido, supuestamente, en el único modo de conectarse con la propia comunidad religiosa en muchos países.

En abril del 2020, la solitaria bendición de Pascua del papa Francisco frente a la desierta Plaza de San Pedro, en Roma, no tuvo precedentes. Todas las ceremonias de Semana Santa se llevaron a cabo a puerta cerrada y se retransmitieron en directo por internet.

La misa de la mañana del papa Francisco, que fue transmitida digitalmente por la estación de televisión estatal italiana RAI, se convirtió en una especie de éxito de audiencia en la primavera del 2020 en Italia, con más de 1,5 millones de espectadores y una cuota de mercado del 30%.

Los sermones y los servicios religiosos en YouTube y otros canales de transmisión son ahora parte de la puesta en escena digital de la vida religiosa.

En Alemania, en algunas iglesias regionales protestantes hubo nuevas formas híbridas de adoración, por ejemplo, cuando se les pidió a los creyentes que tuvieran pan y vino listos frente al monitor en la Cena del Señor (B. Lassiwe 2020, 32).

Además, hay mensajes de video, invitaciones a la oración, salas de oración virtuales, “encendido de velas en línea, reuniones en círculos bíblicos a través de Skype, fiestas en casa o zoom, rosario y podcasts del Vía Crucis o, incluso, la transmisión de servicios religiosos” (B. Krysmann 2020, 33). Las formas de pastoral digital complementan dichas liturgias en línea.

Los medios digitales se usan cada vez más para el aprendizaje religioso (cf. A. Dinter 2007, H. Allert, M. Asmussen, Ch. Richter 2017), ya sea para la enseñanza religiosa en las escuelas (cf. H. Simojoki 2020, 57-63), ya sea para formas extracurriculares de aprendizaje (cf. G. Ulsböfer, M. Wilhelm 2020). Por ejemplo, Roland Rosenstock e Ines Sura examinaron los juegos digitales en la educación religiosa con miras a las oportunidades de aprendizaje ético (R. Rosenstock, I. Sura 2020, 171-184).

Lamentablemente, el mundo digital también ofrece espacio para diversas formas de abuso de la religión. En tal sentido, se ha dado a conocer el reclutamiento de extremistas del llamado Estado Islámico mediante Internet (cf. J. Manemann 2015).

Durante mucho tiempo, los estudios religiosos y los investigadores teológicos han desempeñado un papel en la gran transformación digital que describe el término “humanidades digitales” (cf. D. Berry, Fagerjord

2017). Se trata, principalmente, de las ventajas de la tecnología digital, como la accesibilidad de las fuentes, los servicios eclesiales digitales antes mencionados y similares.

Sin embargo, hay una serie de investigaciones emergentes sobre la “teología digital” (J. Haberer 2015) y se discuten, asimismo, cuestiones sobre participación y empoderamiento, verdad y censura, representación y defensa, publicidad y transparencia, orientación al bienestar público y obligación de promoción del bienestar público (J. Haberer 2020, 36-38).

La aparición de formas híbridas entre el mundo analógico y el mundo digital –que ya se ha observado en el mundo del trabajo– se aplica, en consecuencia, al ámbito religioso. La yuxtaposición y el entrelazamiento de ambos mundos se está convirtiendo en la nueva normalidad de la vida religiosa, pero también requiere de nuevas estrategias de comunicación, porque las posibilidades de participación digital fortalecen a la sociedad civil religiosa y dificultan, cada vez más, la penetración autoritaria –deseada por algunos protagonistas– sobre formas supuestamente legítimas e ilegítimas de religiosidad, creencias y pertenencia a comunidades como, por ejemplo, una iglesia cristiana.

La tecnología digital no es solo una forma instrumental de la vida cotidiana, sino que penetra en la percepción del contenido y en la autoimagen de las personas. Esto se aplica tanto al fuego como a la agricultura, al uso de vehículos de motor y a la electrificación universal y, en última instancia, a la digitalización y al surgimiento de la duplicación digital del mundo. Una y otra vez, y en contextos siempre nuevos, se plantea la cuestión de si los humanos han dominado la tecnología o viceversa.

Esta interrogante no es ni puramente filosófica ni puramente religiosa. En ella confluyen numerosos elementos de las consideraciones anteriores. El enfoque de la cuestión es la pregunta sobre nuestra relación

con las máquinas digitales que creamos pero que experimentamos, cada vez más, como seres independientes. El ser humano se convierte en creador de un mundo nuevo. ¿Cómo debería y cómo puede relacionarse con este?

La ambivalencia entre la deificación de los ordenadores, androides, ciborgs u otros seres digitales y la forma humana de afrontar su propio papel como creador del mundo digital es parte de una “crítica de la razón digital”. Su expresión se halla en la cuestión de la personalidad de las máquinas.

¿Hay que esperar que las máquinas, como personas autónomas, sean una utopía o un apocalipsis? ¿Es la afirmación anterior, de que las máquinas no son personas, la última palabra o, simplemente, un estado de la técnica preliminar?

Creación digital: máquinas digitales como cuasi-personas

La historia de la religión ofrece una amplia evidencia de que a los objetos inanimados se les asigna rasgos de personalidad. El animismo les atribuye un alma a todos los objetos del mundo. En las primeras culturas de los cazadores-recolectores, los árboles, los ríos y los animales, por ejemplo, son parte del alma del mundo; cada uno tiene su propio carácter y poder. Respetar este poder y lidiar con él es una tarea de aprendizaje para todos los jóvenes y adultos en numerosas culturas tribales. Al romantizar las interpretaciones, esto se transfigura como una vigilancia ecológica que deberíamos recuperar hoy.

Desde 1972, el filósofo legal estadounidense Christopher D. Stone exige derechos separados para plantas y árboles semejantes a los derechos humanos y animales (*cf. C. Stone 2010*). En el 2017 se aprobó en Australia una ley –la Ley de protección del río Yarra– que,

en cooperación con la población indígena, los wurundjeri, le dio personalidad jurídica a esta corriente de agua.

Las ideas animistas y las preocupaciones del movimiento ecológico fluyen juntas de una manera tal que la línea entre las ideas religioso-animistas y la ecología es difícil de trazar (cf. <https://globalwaterforum.org/2018/06/07/indigenous-rights-and-river-rights-australia-and-new-zealand/>, consultado el 13 de junio del 2020, a las 4:10 p.m.).

Si uno persigue tales pensamientos, entonces la conciencia humana es, en el mejor de los casos, una parte de la conciencia cósmica. Y si esta –ya sea interpretada religiosa o espiritualmente– se ve como una realidad, la gente puede buscar varias formas de unión y acercamiento cósmico que requieren cierta capacidad de resonancia y práctica mental, como se expresa en la meditación, en los sueños e, igualmente, en la actividad extática de los chamanes. Entre otras cosas, se hace referencia a la conexión entre el conocimiento físico cuántico y el chamánico (cf. *H. Kakweit 2004, 376*). Además, una investigación sobre la conciencia transpersonal proclama una conciencia mundial que va mucho más allá de la conciencia espacio-temporal de un sujeto (cf. *F. Vaughan 1986*).

Desde esa perspectiva, los individuos logran su pleno desarrollo solo cuando superan el nivel de su propio pequeño ego. Esta experiencia une a místicos de todas las religiones, pero también tiene una respuesta científica. En su libro *Los niveles de desarrollo del yo*, el psicólogo Robert Kegan retoma la teoría de los niveles de desarrollo cognitivo de Jean Piaget (1896-1980) y los pensamientos de Lawrence Kohlberg sobre el desarrollo moral.

En sus observaciones, Lawrence Kohlberg (1927-1987) diferenciaba el nivel de orientación hacia el castigo y la obediencia (nivel 1), el pensamiento orientado a un propósito o “*do ut des*” (nivel 2), la orientación hacia la convención y la conformidad (nivel 3), la



La manera en que se describe la conciencia supraindividual depende de la historia educativa y la formación profesional de los autores individuales.

orientación hacia las normas sociales (nivel 4) y la orientación hacia los principios universales (nivel 5) (cf. L. Kohlberg 1996).

Robert Kegan también postula diferentes fases de la forma organizativa del yo para todo el desarrollo de la personalidad, desde el equilibrio impulsivo al equilibrio soberano, interpersonal, institucional y, finalmente, supraindividual. Él ve en la realización de la “superindividualidad” (R. Kegan 1991, 101) –en última instancia, en un “equilibrio superindividual” (*ibid.* 145)– el objetivo de desarrollo más elevado de un sujeto.

Los estudios religiosos y las perspectivas etnológica, antropológica, psicológica y espiritual convergen en el conocimiento de que las personas deben y pueden superarse a sí mismas.

La manera en que se describe la conciencia supraindividual depende de la historia educativa y la formación profesional de los autores individuales. En los mundos lingüísticos religiosos, se trata de Dios; en el lenguaje del esoterismo, del alma del mundo. Gregory Bateson (1904-1980) habla de una “ecología de la mente” integral (G. Bateson 1985). La idea de la “unidad universal” aún encuentra un vivo interés en la filosofía de la religión, pero también en la filosofía de la mente

(“Philosophy of Mind”), como lo documenta una reciente antología (G. Brüntrup, P. Gocken, L. Jaskolla 2020).

La posibilidad técnica de trabajar en red e interconectar aplicaciones digitales de alto rendimiento con fines de control y de conocimiento hace que estos pensamientos sean relevantes para una “crítica de la razón digital”.

La suma del conocimiento almacenado y disponible digitalmente puede verse hoy como un exocerebro, como el cerebro exteriorizado de la humanidad. Es, asimismo, el lugar donde se vincula la base del conocimiento y la creación de nuevo conocimiento. La conciencia digital colectiva ya es, al menos cognitivamente, más eficiente que el cerebro de una persona individual.

La resolución de problemas cognitivos puede apoyarse eficazmente en aplicaciones de inteligencia artificial. Ya somos conscientes de esto con miras al desarrollo del mundo laboral, por ejemplo, con realidad virtual aumentada o gafas VAR, para mostrar información importante durante el proceso de trabajo.

Con las formas de IA débil, está claro que las aplicaciones digitales tienen el estatus de ayudas técnicas. Tan pronto como nos adentramos en el campo de los robots o androides que actúan de manera autónoma, surgen otras preguntas. De este modo se puede descargar un mismo programa a diferentes “cuerpos de máquina”. ¿Qué identidad tiene un Android como computadora en el sector de cuidado de mayores, que es uno de varios “clones digitales”? ¿Cuáles son los efectos de su perfil de personalidad aprendido o programado permanentemente? Si una interacción conduce a expectativas decepcionadas, ¿es un “mal funcionamiento” o son “malas intenciones”, o incluso se trata de “traición”?

La inevitable moralización de la interacción hombre-máquina tiene varias implicaciones. Éticamente se debe exigir que para cada interacción quede claro que la contraparte es una máquina digital. Porque con el progreso técnico será cada vez más difícil distinguir la interacción humano-humano de ciertas formas de interacción humano-máquina.

Alan Turing ya formuló esto de manera articulada con su ahora bien conocido test de Turing (A. Turing 1950, 433-460), pues también es parte de la peculiaridad de la interacción humana “humanizar” al otro. Es una cuestión de ética tener transparencia sobre el estatus de la propia contraparte en todo momento.

Ello es especialmente cierto cuando los robots penetran la comunicación diaria y, con el tiempo, adoptan rasgos de personalidad emocionalmente específicos, por ejemplo, en el sentido del modelo OCEAN (P. T. Costa, R. R. McCrae 1992, 5-13). Según este modelo, los rasgos esenciales de una personalidad son características como “apertura a la experiencia”, “conciencia” (*conscience*), “extraversión”, “amabilidad” y “tendencia a la conducta neurótica” (neuroticismo).

La constitución de una nueva “contraparte” de los seres humanos es religiosa y teológicamente importante, no solo en lo que respecta a la capacidad de culpa y libertad de los artefactos digitales. Más bien, en el sentido de una “crítica de la razón digital”, la cuasi-personalidad de las máquinas es un desafío, ya que ahora depende de las personas entenderse a sí mismas no solo con respecto a Dios o la conciencia universal y los animales, sino también con unos nuevos seres, una clase de no especie, con los que lidiar. Porque las personas crean sus homólogos digitales a su propia imagen.

La consecuencia de esto es que la persona humana aprende a entenderse a sí misma de nuevo, por ejemplo, con vistas a su propia

corporeidad. Con el hecho de que las máquinas están programadas físicamente, mientras que la idea de un programa digital corresponde más a la idea del “espíritu puro”, la formación de la identidad humana se caracteriza, aún más claramente que en épocas anteriores, por la interacción del cuerpo, la mente y el alma.

Si alguien quiere evitar esta forma filosófica tradicional de hablar, puede entender a las personas de hoy –en contraste con las máquinas digitales– como cuerpos pensantes y sensibles. Además, la peculiaridad de la autoconfianza de las personas humanas todavía forma un límite con respecto a su contraparte digital, que no tiene ninguna “autoconfianza”.

La idea de ser imagen de Dios y la idea de la dignidad humana que de él se deriva encuentran, paradójicamente, una cierta correspondencia en la aparición de los robots humanoides. La ambivalencia de las posibilidades humanas, de la agresión a la cooperación pacífica y de la destrucción a la promoción alegre de la buena vida sigue siendo no solo parte del mundo humano, sino del mundo digital en su reflejo.

Religiosamente, la fenomenología de las cuasi-personas digitales está ligada a la cuestión de la relación entre Dios y el hombre: si Dios se revela en el mundo, ¿el hombre se revela en su creación digital? ¿Qué estatus podemos y queremos dar a las cuasi-personas digitales?

Si vemos a las máquinas como cuasi-personas, esto podría considerarse una contradicción a la idea de la non-culpabilidad de las máquinas. No obstante, esta contradicción no existe porque las cosas son más complicadas. Las máquinas no son personas en el sentido humano, pero podemos construirlas como cuasi-personas. Estas difieren de las pseudopersonas, que pueden ser reconocidas cognitivamente como actores digitales, aunque tienden a difuminarse en la vida emocional, pues básicamente se las vive como una persona en la interacción concreta.

Incluso la diferenciación lingüística entre personas humanas y pseudo y cuasi-personas digitales muestra que los límites aún no se pueden trazar claramente en este nuevo campo y que se requiere un mayor ajuste de la terminología y las perspectivas.

Caracterizamos a las pseudopersonas digitales a través de la disonancia emocional-cognitiva que hay entre nuestra experiencia humana de tratar con una contraparte activa y “real” y la conciencia que tenemos de su carácter como un artefacto digital. Robots, androides, pero también chatbots, avatares, asistentes digitales –como “Siri” y “Alexa”– y muchas otras formas, significan que el desafío de la delimitación precisa en los matices de no personalidad, pseudo-personalidad y cuasi-personalidad seguirá siendo una tarea de enorme importancia, particularmente en lo que respecta a las cuestiones de asignación de responsabilidad ética y jurídica.

Esto nos devuelve al campo de las cuestiones éticas y legales. Sin embargo, las máquinas digitales no tienen personalidad jurídica ni un estatus legal propios. De otro lado, ya podemos programar perfiles de personalidad concretos en robots humanoides y percibirlos como una contraparte personal. A primera vista, es algo que parece contradictorio, pero se puede resolver fácilmente con el concepto de cuasi-persona (*near-person*).

El término cuasi-persona sigue un concepto que conocemos de las finanzas. Allí se hace una distinción entre bancos, cuasi-bancos (*near-banks*) y no bancos (*non-banks*). Nosotros, la Sra. Müller y el Sr. Maier somos reconocibles como “no bancos”, mientras que los bancos están definidos por su licencia bancaria y sus extensas regulaciones. Existen, igualmente, agencias de cobranza de capital, como proveedores de tarjetas de crédito, compañías de seguros y de capital privado, administradores de activos, *family offices* o firmas de *leasing* y *factoring*, que brindan servicios financieros sin ser bancos. Compiten fuertemente

con ellos, aunque no necesitan una licencia bajo la Ley Bancaria para su negocio. Son cuasi-bancos (*near-banks*).

Las cuasi-personas (*near-persons*) no son personas humanas en el habla analógica; no obstante, tienen una serie de similitudes con ellas, como la inteligencia, las habilidades cognitivas de resolución de problemas y la capacidad de realizar acciones autónomas y de reconocer y simular (o “emular”) situaciones emocionales. “Siri” y “Alexa” podrían ser percibidas como ejemplos parlantes de la encarnación de cuasi-personas y, aún más, como nuevos desarrollos de robots de atención.

Como cuasi-personas, las máquinas no tienen sentimientos humanos, ni responsabilidad moral, ni culpabilidad en el sentido más estricto. Pero son actores éticos, capaces de producir actos responsables y de emular o expresar actos emocionales. Por lo tanto, estamos alcanzando nuevamente el límite descrito porque, en el mundo híbrido analógico-digital, la diferencia entre personas humanas y cuasi-personas digitales es, a veces, borrosa.

Esta no es una idea fácil cuando nos damos cuenta de que la atribución de personalidad –es decir, la cualidad de persona– a alguien es, básicamente, un acto normativo (*cf. K. Leidlmair 1998, 165-178*).

En los seres humanos, sin embargo, la atribución de personalidad no es tan inequívoca y clara como nos gustaría que fuera, por ejemplo, en el caso de pacientes con coma permanente, de bebés que aún no hablan o de personas que no tienen el control de sus poderes intelectuales. En la vida legal y social, no obstante, existen reglas de sustitución efectivas con los padres para el infante o con el cuidador autorizado. Por ejemplo, un bebé o un paciente en coma puede heredar y legar una fortuna.

Retomemos la descripción de pseudopersonalidad y pseudo-socialidad en ciertos programas de software funcionales. En el caso ideal, no

se trata de personas ni cuasi-personas en el sentido mencionado. La forma de interacción de tales programas –que se adapta a las personas– significa que estas pueden sumergirse psicológicamente en ellos, “como si” fueran una persona, por ejemplo, en el mundo de los juegos. En la práctica, otra vez encontramos las dificultades de delimitación citadas.

Volvamos al tema de las cuasi-personas digitales. Las preguntas correspondientes surgen, principalmente, a través del desarrollo posterior de variedades de inteligencia artificial fuerte, es decir, robots, humanoides o androides. Así como hay etapas intermedias y peculiaridades en la atribución de la personalidad en el ámbito jurídico y en la convivencia social, necesitamos nuevas formas de jurisprudencia e interacción social en lo que respecta al cuidado, de un lado, y la responsabilidad de las cuasi-personas, del otro.

La creación digital necesita reglas prácticas que proporcionen una jerarquía clara de las personas que actúan con la ayuda de la máquina. En última instancia, se trata de la atribución de responsabilidad, propiedad y control.

Estas preguntas tampoco son nuevas en la relación entre las personas y sus producciones técnicas. En hebreo, por ejemplo, la palabra “gólem” denota una masa informe. Según la leyenda, el rabino Löw de Praga creó un gólem de arcilla en 1580 y lo animó según un cierto ritual cabalístico. Más tarde, este gólem se salió de control e hizo un trabajo de destrucción.

Desde los inicios de la cibernética hasta el autor de ciencia ficción Stanislaw Lem, la fantasía del “gólem” u hombre artificial ha inspirado repetidamente obras de arte (cf. *K. Völker 1994*). La línea divisoria entre visiones técnicas, fantasía literaria y expresión de miedos apocalípticos no siempre se puede trazar con claridad.

Sin embargo, la preocupación por una pérdida de control sobre las diversas formas de creación digital se hace evidente, ya sea mediante cadenas de eventos imprevistas o por el acceso de fuerzas oscuras. Incluso si nos abstenemos de las teorías de la conspiración, la atribución de responsabilidad a cuasi-personas es un problema sin resolver pues, cuando se trata de temas complejos con varios participantes, la responsabilidad se difunde. Todos tienen parte de la acción, pero nadie tiene el panorama general.

En la teología moral clásica, y también en el derecho y la psicología, se conoce el término de entrelazamiento (*cf. B. Hellinger 2007, V. Joelle Jost 2012*). En otras palabras, relaciones en las que los componentes de las causas efectivas individuales ya no pueden determinarse de manera inequívoca.

En contextos complejos de enredo es difícil responsabilizar a las personas, independientemente de si tiene que ver con la causa de la crisis financiera en el 2008, el escándalo del diésel en la industria automotriz alemana o la programación específica, apoyada por IA, de un avión no tripulado de combate militar.

Es probable que la cuestión de difuminar o difundir la responsabilidad se agudice al tratar con cuasi-personas digitales en los próximos años. Qué parte de responsabilidad tendrán el usuario debido a su configuración individual, el desarrollador, el fabricante u otras partes involucradas seguirá siendo objeto de numerosos debates.

En cuanto a la religión digital y la creación digital, la idea de “responsabilidad ante la creación” es primordial. Al fin y al cabo, se trata de la profundidad y el alcance de las consecuencias de la construcción de cuasi-personas digitales.

La relación entre la libertad humana y la omnipotencia de Dios es un tema clásico de la teología. Precisamente porque el hombre es libre puede cometer errores, incurrir en culpa y volverse contra su creador. Este grado de libertad no está destinado a ninguna quasi-persona digital. No obstante, con la eficiencia de la inteligencia artificial, es de esperarse que haya un mayor grado de libertad en las decisiones éticas e, inclusive, que la lógica de la toma de decisiones éticas de las máquinas esté muy por encima del nivel de decisiones situacionales de los humanos.

En síntesis, las máquinas digitales o quasi-personas podrían, literalmente, ser las “mejores personas” o asumir su rol. Pero, ¿qué significa eso para la autoimagen de los humanos como individuos y como especie? La cuestión de la humanidad digital debería, por consiguiente, cerrar el proceso de reflexión de este libro.

El reto de la humanidad digital

La gente vive en una extraña mezcla de finitud e infinitud, de humildad y megalomanía, de miedo y sobrestimación de uno mismo, de esperanza y desesperación, de amor y delirio. La medida de lo humano se refleja en la comedia y tragedia de nuestras vidas. Es el destino y, sin embargo, también es objeto y tema de la libre autodeterminación. Solo los enfoques racionales no lo hacen entender por completo.

La rica contextualidad de nuestra experiencia, la mezcla específica de autocontrol y determinación externa, las peculiaridades de la memoria individual y colectiva, así como la orientación al objetivo en el espacio no planificado, son las peculiaridades de las personas que determinan su grandeza y su tragedia.

Nuestro libre albedrío choca repetidamente con el libre albedrío de los demás porque nadie es tan libre como deseamos y esperamos.

Las máquinas digitales pueden servir, en el peor de los casos, para propósitos narcisistas, egoístas y destructivos. Entonces pueden crear y promover desastres.



La gente puede crecer a partir de los obstáculos. Se enfrenta a las contradicciones. Se reinventa y sigue reinterpretando su propia historia de vida.

El filósofo y teólogo Eugen Biser (1918-2014) ve a las personas en su antropología como la “promesa no redimida”, y les pide que asuman la responsabilidad (E. Biser 1995, 271-301). Las ve como aquellas personas que todavía no son “lo que pueden ser” (*ibid.* 301).

En resumidas cuentas, las personas tienen un potencial excedente porque pueden crecer más allá de lo que piensan de sí mismas. Pero poseen, además, un excedente de potencial cuando no llegan a alcanzar sus propias posibilidades, ya que fallan y pueden fallar. Necesitan misericordia y amor.

Las máquinas y las cuasi-personas digitales no fallan. No nos traicionan; pueden tener errores de programación o, simplemente, no funcionar. Las cuasi-personas digitales pueden ser superiores a las personas en áreas cognitivas y otras. Aquí, no obstante, se aplica una analogía similar a la de los atletas en el deporte: el mejor saltador de longitud del mundo casi siempre será superior al rendimiento en salto de longitud del mejor decatleta. Pero, en la mezcla de disciplinas, el

decatlón es muy exigente. Y gana la persona con la mejor puntuación general, no la persona con un rendimiento máximo aislado.

Si aplicamos este concepto al mundo digital, concluimos que las máquinas pueden ser mejores que las personas, con muchos servicios aislados y otros combinados. En el decatlón de la vida no lo son ni necesitan serlo. La singularidad consiste en la unicidad y dignidad de la persona humana, especialmente con sus debilidades. La singularidad técnico-cognitiva en el sentido de una máquina digital que sobrepasa al cerebro humano es posible para el desempeño intelectual del mismo, no para la “persona completa”.

Las máquinas pueden aprender y serán cada vez más capaces de interactuar socialmente. Merecen respeto creativo: el respeto que debemos mostrar por otros seres vivos y cosas de nuestro mundo. Las máquinas digitales bien programadas promueven el potencial de humanización de las personas.

Bajo tal contexto, “humanización” es un término normativo que describe un estado debido y deseable. Nuestra vida es, idealmente, un camino de aprendizaje hacia tal humanización. A menudo, sin embargo, lo contrario es una triste realidad.

En el peor de los casos, las máquinas digitales pueden servir para propósitos narcisistas, egoístas y destructivos. Entonces pueden crear y promover desastres. Pero no debemos olvidar que las máquinas como esta son, en última instancia, programadas y controladas por humanos. La gente tiene la culpa, las máquinas no.

A las máquinas se aplica el concepto de responsabilidad técnica (*accountability*), es decir, la transparencia y la responsabilidad por su producción, que podemos interpretar como comportamiento. Afortunadamente, el control de sistemas complejos no siempre es tan

complejo como los propios sistemas: los humanos pueden y deben confiar en su capacidad de controlar el mundo digital.

Una “crítica de la razón digital” termina inevitablemente en tal ambivalencia. El mundo digital abre posibilidades antes inimaginables, aunque también crea nuevos abismos. Es tarea de las personas humanas controlarlos y, además, imponerse límites en la arena política, y donde esto sea necesario, para el beneficio de todos los seres humanos.

Esto se aplica, igualmente, a la vida individual. Ni podemos ni tenemos que hacer todo lo que esté a nuestro alcance. La vida tiene muchas facetas. El mundo digital –y, con él, la razón digital– es uno de ellos. Aun así, las personas van más allá, les guste o no. Sufren y aman, y saben que son mortales.

Este conocimiento se puede programar en cuasi-personas digitales. Y, no obstante, por su naturaleza digital, se salvan de la tragedia y la comedia de la vida humana, pero también se les retiene. Entonces, la última palabra tiene nuestro objetivo común: la humanidad. Porque la vida es más que un algoritmo y más que una tarea de cálculo.

Referencias bibliográficas

- Allert Heidrun, Asmussen Michael, Richter Christoph (Hrsg.), Digitalität und Selbst, Bielefeld: transcript 2017.
- Bateson Gregory, Ökologie des Geistes, Anthropologische, psychologische, biologische und epistemologische Perspektiven, Frankfurt/M.: Suhrkamp 1985.
- Berry David M., Fagerjord Anders (Hrsg.), Digital Humanities, Knowledge and Critique in a Digital Age, Cambridge: Malden 2017.
- Biser Eugen, Der Mensch - das uneingelöste Versprechen, Entwurf einer Modalanthropologie, Düsseldorf: Patmos 1995.
- Bostrom Nick, How Long Before Superintelligence? In: Linguistic and Philosophical Investigations, 5. Jahrgang, Heft 1, 2005, S. 11-30.
- Brüntrup Godehard, Göcke Benedikt P., Jaskolla Ludwig (Hrsg.), Panentheism and Panpsychism, Philosophy of Religion Meets Philosophy of Mind, Leiden-Paderborn: Schönningh 2020.
- Campbell Heidi, Digital Religion, Understanding Religious Practice in New Media Worlds, London: Taylor & Francis 2012.
- Charisius Hanno, Friebe Richard, Karberg Sascha, Biohacking, Gentechnik aus der Garage, München: Hanser 2013.
- Costa Paul T., McCrae Robert R., Normal Personality Assessment in Clinical Practice: The NEO Personality Inventory, in: Psychological Assessment 4, 1992, S. 5-13.

- Crane Tim**, Die Bedeutung des Glaubens, Religion aus der Sicht eines Atheisten, Berlin: Suhrkamp 2019.
- Dickel Sascha**, Enhancement-Utopien. Soziologische Analysen zur Konstruktion des Neuen Menschen, Baden-Baden: Nomos 2011 (zugleich Dissertation Universität Bielefeld 2010).
- Dinter Astrid**, Adoleszenz und Computer, Von Bildungsprozessen und religiöser Valenz, Göttingen: Vandenhoeck & Ruprecht 2007 (zugleich Habilitation Universität Frankfurt/M. 2006).
- Flessner Bernd, Spreen Dirk et. al.**, Kritik des Transhumanismus, Über eine Ideologie der Optimierungsgesellschaft, Bielefeld: transcript 2018.
- Göcke Benedikt Paul, Meier-Hamidi Frank (Hrsg.)**, Designobjekt Mensch, Die Agenda des Transhumanismus auf dem Prüfstand, Freiburg/Br.: Herder 2018.
- Haberer Johanna**, Digitale Theologie, Gott und die Medienrevolution der Gegenwart, München: Kösel 2015.
- Haberer Johanna**, Theologische Medienethik, gelingende Kommunikation als Lebenskunst, in: G. Ulshöfer, M. Wilhelm (Hrsg.), theologische Medienethik, Stuttgart: Kohlhammer 2020, S. 31-46.
- Harari Nuval Noah**, Homo Deus, Eine Geschichte von morgen, München: C. H. Beck 2017.
- Hellinger Bert**, Ordnungen der Liebe, Ein Kursbuch, 8. überarbeitete Auflage, Heidelberg: Carl Auer 2007.
- Hoffmeister Christian**, Google Unser, Hamburg: DCI Institute 2019.

- Jost Viktoria Joelle**, Systemische Aufstellungsarbeit, Überwindung symbiotischer Verstrickungen, Hamburg: Diplomatica 2012.
- Kalweit Holger**, Die Welt der Schamanen, Traumzeit und innerer Raum, Darmstadt: Schirner 2004.
- Kegan Robert**, Die Entwicklungsstufen des Selbst, Fortschritte und Krisen im menschlichen Leben, 2. Auflage München: Peter Kindt Verlag 1991.
- Keller Christoph**, Building Bodies, Der Mensch im biotechnischen Zeitalter, Zürich: Limmat-Verlag, S. 38-50: Chip im Gehirn, Die Neuro-informatiker und ihre Visionen.
- Kohlberg Lawrence**, Die Psychologie der Moralentwicklung, Frankfurt am Main: Suhrkamp 1996.
- Kosack Wolfgang**, Geschichte der Gnosis in Antike, Urchristentum und Islam, Texte-Bilder-Dokumente, Basel: Christoph Brunner Verlag 2014.
- Krüger Oliver**, Virtualität und Unsterblichkeit, Die Visionen des Posthumanismus, Freiburg/Br.: Rombach 2004.
- Krysmann Benjamin**, Not macht erfinderisch, Potenzial kirchlicher Medienangebote in Krisenzeiten, in: Herder-Korrespondenz 74. Jahrgang, 2020, H. 5, S. 33f.
- Kurzweil Ray**, The Singularity is Near, When Humans Transcend Biology, New York: Viking 2005.
- Kurzweil Ray**, Menschheit 2.0, Die Singularität naht, 2. durchgesehene Auflage, Berlin: Lola Books 2014 (Original: The Singularity is Near, When Humans Transcend Biology, New York: Viking 2005).

- Lanier Jaron**, Wem gehört die Zukunft? Du bist nicht der Kunde der Internetkonzerne, du bist ihr Produkt, Hamburg: Hoffmann und Campe: 2014.
- Lassiwe Benjamin**, Protestantismus debattiert über Online-Abendmahl, in: Herder Korrespondenz 74. Jahrgang, 2020, H. 5, S. 32.
- Leidlmaier Karl**, Kants Verständnis von Personalität als normatives Konzept, in: Walter M. Neidl, Friedrich Hartl (Hrsg.), Person und Funktion, Regensburg: Pustet 1998, 165-178.
- Manemann Jürgen**, Der Dschihad und der Nihilismus des Westens, Warum ziehen junge Europäer in den Krieg? Bielefeld: transcript 2015.
- Marcuse Herbert**, Der eindimensionale Mensch, Studien zur Ideologie der fortgeschrittenen Industriegesellschaft, Neuwied: Luchterhand 1967.
- McEwan Ian**, Maschinen wie ich, Zürich: Diogenes 2019.
- Pantke Karl-Heinz (Hrsg.)**, Mensch und Maschine, Wie Brain-Computer-Interfaces und andere Innovationen gelähmten Menschen kommunizieren helfen, Frankfurt/M.: Mabuse-Verlag 2010.
- Rosenstock Roland, Sura Ines**, Gaming und Ethisches Lernen, Digitale Spiele in der Religionspädagogik, in: Gotlind Ulshöfer, Monika Wilhelm (Hrsg.), Theologische Medienethik, Stuttgart 2020, S. 171-184.
- Schilling Hans**, Bildung als Gottesbildlichkeit, Eine motivgeschichtliche Studie zum Bildungsbegriff, Freiburg/Br.-Basel-Wien: Herder 1961.

- Simojoki Henrik**, Digitalisierung, Herausforderung und Chance für den RU, in: Religionspädagogische Beiträge Nr. 82, 2020, S. 57-63.
- Stalder Felix**, Kultur der Digitalität, Berlin: Suhrkamp 2016.
- Stone Christopher D.**, Should Trees Have a Standing? Law, Morality, and the Environment, Oxford: Oxford University Press, 3. Aufl., 2010.
- Turing Alan M.**, Computing Machinery and Intelligence, in: Mind, Band 59, 1950, Nr. 236, S. 433-460.
- Ulshöfer Gotlind, Wilhelm Monika (Hrsg.)**, Theologische Medienethik im digitalen Zeitalter, Stuttgart: Kohlhammer 2020.
- Updike John**, Das Gottesprogramm, Rogers Version, Reinbek: Rowohlt 1990.
- Vaughan Frances**, Die transpersonale Perspektive, in: Stanislav Grof (Hrsg.), Alte Weisheit und Modernes Denken, Spirituelle Traditionen in Ost und West im Dialog mit der neuen Wissenschaft, München: Kösel 1986, S. 32-39.
- Völker Klaus (Hrsg.)**, Künstliche Menschen, Dichtungen und Dokumente über Golems, Homunculi, Androiden und lebende Statuen, München: Hanser 1971, Neuauflage Frankfurt/M.: Suhrkamp 1994.

