Hacia una ética relacional de la responsabilidad

Un marco ético para el diseño tecnológico en la teoría de la mente extendida

Autor: Óscar Bodí Pons
Directora: Cristina Llorca García

Enero de 2025
TFG del Grado de Humanidades
Estudios Interdisciplinarios

Universitat Oberta de Catalunya

Índice

Resumen

Este trabajo aborda el desafío ético de cómo ciertas tecnologías, al integrarse como extensiones de nuestra mente, transforman profundamente nuestra autonomía, privacidad y subjetividad, desdibujando los límites entre lo humano y lo artificial. Si bien estas herramientas pueden ampliar nuestras capacidades cognitivas, también plantean riesgos como la manipulación, la dependencia y la erosión de la privacidad mental. Uno de los principales retos es definir quién debe asumir la responsabilidad ética en este contexto de interdependencia humano-tecnológica. Para responder a estas tensiones, se propone una ética relacional de la responsabilidad fundamentada en la Teoría de la Mente Extendida. Este marco, basado en los principios de claridad y autonomía, responsabilidad interdependiente y justicia, y prospectiva ética y beneficencia, busca no solo guiar el diseño de tecnologías que respeten la dignidad humana y fortalezcan la capacidad crítica, sino también distribuir equitativamente las responsabilidades entre los distintos actores implicados.

Palabras clave: ética relacional; tecnologías mente extendida; riesgos cognitivos en tecnología; autonomía relacional; privacidad mental.

1. Introducción

1.1. Presentación

> ¿Debemos elegir entre el Apocalipsis y el futuro radiante?
> Más bien, creo que hay que retroceder un poco tratando de indagar
> de dónde pueden proceder sentimientos tan contradictorios.
> (Latour, 2012)

Imagina un asistente digital que, al recordarte cada cita y sugerir rutas en tu ciudad, se convierte en una parte inseparable de tu proceso de decisión diaria. Ahora, reflexiona: ¿qué ocurre si este asistente comienza a priorizar ciertas rutas o actividades, no por optimización, sino por intereses comerciales de la empresa que lo desarrolló? Este escenario, que puede parecer futurista, ya refleja las tensiones éticas y prácticas de la integración humano-tecnológica en nuestra cotidianidad.

En un momento donde la tecnología redefine nuestra forma de pensar, decidir y actuar, este trabajo se enmarca en la Teoría de la Mente Extendida (TME) para examinar cómo dispositivos digitales e inteligencia artificial actúan no solo como herramientas, sino como extensiones cognitivas que transforman nuestras capacidades y decisiones. Desde esta perspectiva, surge una cuestión clave: ¿cómo podemos garantizar que estas tecnologías respeten nuestra autonomía, privacidad y dignidad, en lugar de condicionarlas o erosionarlas? La respuesta requiere no sólo un análisis crítico de los retos éticos, sino también la construcción de un marco ético que oriente su diseño hacia un futuro más justo.

1.2. Tema central y objetivos del estudio

Las tecnologías digitales y la inteligencia artificial han dejado de ser simples herramientas auxiliares para convertirse en elementos integrados profundamente en la cognición humana. Según la TME, propuesta por Clark y Chalmers (1998), las capacidades cognitivas humanas no se limitan al cerebro, sino que se amplían a través de la interacción con dispositivos tecnológicos. El tema central de este trabajo es explorar cómo las tecnologías actúan como extensiones de la mente y evaluar las implicaciones éticas de esta integración, particularmente en aspectos como la autonomía, la privacidad y la subjetividad.

El presente trabajo tiene como objetivo principal evaluar los desafíos éticos derivados de la TME y proponer un marco ético con enfoque en la responsabilidad para guiar el diseño y uso de las tecnologías. A partir de este objetivo, se derivan los siguientes objetivos específicos:

- Analizar los fundamentos teóricos de la TME.
- Examinar el impacto de las tecnologías desde una perspectiva teórica y crítica.
- Evaluar la aplicabilidad y limitaciones de los marcos éticos tradicionales.
- Proponer un marco ético para las tecnologías en la TME.

A través de este análisis, este trabajo busca no solo enriquecer el debate ético sobre la relación entre tecnología y cognición, sino también aportar herramientas que permitan diseñar tecnologías que respeten el bienestar y dignidad de los usuarios.

1.3. Estado de la cuestión y marco teórico

El presente trabajo se construye desde un enfoque interdisciplinario que combina las perspectivas de la TME, la ética aplicada, la neurociencia, la filosofía de la mente y el diseño tecnológico. Estas disciplinas interactúan para ofrecer un marco integral que permita analizar las implicaciones éticas y cognitivas de las tecnologías en la relación humano-tecnológica. "The Extended Mind" de Andy Clark y David Chalmers (1998) es clave para este marco teórico, ya que introduce la idea de que la cognición humana trasciende el cerebro y se extiende a herramientas y entornos tecnológicos que interactúan activamente con nuestros procesos mentales. Esta teoría redefine los límites de la mente y destaca cómo las tecnologías se convierten en parte de nuestra subjetividad y agencia.

La idea de cognición distribuida, desarrollada por Edwin Hutchins (2000) en "Cognition in the Wild" y Francisco Broncano (2006) en "Sujeto y subjetividad en la mente extensa", amplía la comprensión de cómo los artefactos tecnológicos afectan la subjetividad y la cognición humana. Por su parte, David Kirsh, en su obra "Explaining Artifact Evolution" (2006), analiza cómo los artefactos tecnológicos evolucionan y se integran en los procesos cognitivos humanos, desempeñando un papel clave en la externalización de funciones como la memoria y la planificación. Estas perspectivas ofrecen una base conceptual crítica para el desarrollo de una ética que busca abordar estas interacciones complejas.

En el ámbito de la ética, Adela Cortina (2024), en "Ética o ideología de la inteligencia artificial", analiza los riesgos de la tecnologización moderna, subrayando la necesidad de principios de justicia y dignidad humana. Hans Jonas (1995), en "El principio de

responsabilidad", plantea una ética orientada al futuro que inspira la evaluación de riesgos tecnológicos en este trabajo. Carol Gilligan (2013), en "La ética del cuidado", aporta una ética basada en el cuidado y el reconocimiento del Otro, integrando la empatía y respeto en el marco ético propuesto. Estas teorías aportan un enfoque dirigido al cuidado en las relaciones humanas, no solamente presente, sino orientado al futuro.

En el ámbito de la autonomía, Jennifer Nedelsky (2011), en "Law's Relations", redefine la autonomía desde una perspectiva relacional, sentando las bases para una ética que priorice las relaciones humanas y su papel en la configuración de la autonomía. Por su parte, Yifat Braudo-Bahat (2017) desarrolla una conceptualización interdependiente en la relación humano-tecnológica, destacando la necesidad de una autonomía relacional. Verbeek (2011), en "Moralizing Technology", complementa estas perspectivas al analizar cómo los artefactos tecnológicos no solo median las relaciones humanas, sino también modelan la moralidad y las decisiones. Estas contribuciones articulan la tecnología y las relaciones humanas en un marco orientado hacia la relacionalidad e interdependencia.

En relación con los riesgos para la privacidad mental y la manipulación, John Burrell (2016) aborda la opacidad de los algoritmos en su obra "How the machine ‘thinks’", identificando cómo estos sistemas pueden influir en nuestras decisiones sin nuestro conocimiento. Clowes, Smart y Heersmink (2024) amplían esta discusión en "The ethics of the extended mind", donde exploran la importancia de proteger la privacidad mental como un componente esencial de la autonomía personal. Michael Blitz (2010), en "Freedom of thought for the extended mind", complementa esta perspectiva al enfatizar que la libertad de pensamiento en un contexto tecnológico requiere garantizar que las tecnologías cognitivas no invadan ni manipulen los procesos internos de los usuarios. Así mismo, en "Data, metadata, mental data?", Palermos (2023) aborda los desafíos para la privacidad mental en un mundo de datos, destacando los riesgos emergentes relacionados con la vigilancia algorítmica. Estas reflexiones priorizan la protección de la privacidad mental como un pilar de la dignidad y la justicia en entornos interconectados.

La idea de agencia extendida cuestiona los límites tradicionales de la acción humana, destacando cómo los artefactos tecnológicos contribuyen activamente a redes de interacción que transforman no solo la responsabilidad, sino también las estructuras morales en las que estas operan. Bruno Latour, en su Teoría del Actor-Red (Argemí y Tirado, s.f.), plantea que tanto humanos como objetos tecnológicos actúan como actores equivalentes en la configuración de redes sociales y tecnológicas. Hanson (2009), en "Beyond the skin bag", profundiza en esta idea al explorar la responsabilidad compartida entre humanos y

dispositivos tecnológicos. Gunkel (2020), en "Mind the gap: Responsible robotics and the problem of responsibility", subraya las limitaciones actuales en la atribución de responsabilidad dentro de sistemas tecnológicos complejos. Su análisis destaca la necesidad de enfoques éticos que consideren la agencia híbrida y los artefactos tecnológicos como actores clave. Estas perspectivas convergen para fundamentar la necesidad de un enfoque ético que considere las interacciones entre humanos y tecnologías como dinámicas y mutuas, abordando las brechas existentes en la atribución de responsabilidad.

Otros autores también contribuyen de manera significativa al marco teórico. Por ejemplo, Luciano Floridi (2024) en "Ética de la inteligencia artificial" proporciona un marco para abordar los principios y desafíos éticos en el marco de la IA. Daniel Dennett (1996), en "Kinds of Minds", explora los fundamentos de la conciencia y su relación con las tecnologías cognitivas, mientras que Bratman (2000) en "Reflection, planning, and temporally extended agency" analiza cómo las tecnologías influyen en nuestra capacidad de planificación y agencia extendida. Albrechtslund (2007), en "Ethics and technology design", propone un enfoque ético centrado en cómo las tecnologías deben diseñarse para promover la confianza y el bienestar humano. Friedman et al. (2002) introducen el concepto de "Value Sensitive Design", que enfatiza la integración de valores humanos en el proceso de diseño tecnológico.

En conclusión, este marco teórico proporciona una base integral para analizar las tecnologías cognitivas desde una perspectiva interdisciplinaria, conectando teorías clave con los principios de autonomía, privacidad, justicia y dignidad. La ética relacional de la responsabilidad, propuesta en este trabajo, emerge como una respuesta innovadora a las lagunas en los enfoques actuales, ofreciendo un marco que orienta la responsabilidad del diseño tecnológico hacia la preservación de relaciones basadas en la dignidad y la justicia.

1.4. Metodología y técnicas de investigación

El presente trabajo adopta un enfoque teórico-conceptual e interdisciplinario, diseñado para analizar las implicaciones éticas de las tecnologías cognitivas desde la perspectiva de la TME. Este enfoque cualitativo es especialmente adecuado debido a la naturaleza emergente y compleja del tema, que requiere integrar conocimientos provenientes de múltiples disciplinas, como la ética aplicada, la filosofía de la mente, la neurociencia y el diseño tecnológico.

La revisión bibliográfica se realizó mediante un proceso sistemático en bases de datos académicas de acceso abierto y restringido, como Biblioteca UOC, Google Scholar, PhilPapers, ResearchGate e IEEE Xplore. Los criterios de inclusión para las fuentes fueron:

1. Actualidad: Priorizar textos publicados en los últimos 15 años para garantizar la relevancia de las perspectivas, sin excluir obras seminales fundamentales como "The Extended Mind" de Clark y Chalmers (1998).
2. Interdisciplinariedad: Incluir estudios de diversas disciplinas, desde análisis éticos hasta casos empíricos en neurociencia y diseño de tecnología.
3. Idioma: Considerar publicaciones en inglés y español para captar la amplitud del debate académico en torno al tema.
4. Relevancia conceptual: Identificar textos clave relacionados con autonomía, privacidad mental y ética aplicada en el contexto de la TME.

El proceso de búsqueda utilizó palabras clave como *Technology of the Extended Mind*, *Ethics of Technology*, *Autonomy Extended Mind*, *Mental Privacy Extended Mind*, entre otras. Las referencias seleccionadas fueron organizadas y citadas siguiendo el formato APA 7ª edición para garantizar la coherencia en la presentación de las fuentes.

2. Tecnología y extensión de la mente

2.1. La tecnología como extensión cognitiva

Desde sus orígenes, el ser humano no solo ha transformado su entorno mediante herramientas y artefactos, sino que también ha sido transformado por ellos en un proceso continuo y bidireccional. En términos evolutivos, esta interacción ha tenido un impacto tanto a nivel biológico —en la evolución de la especie (filogénesis) al desarrollar habilidades como el bipedismo, la prensión manual o el crecimiento de la corteza cerebral— como en el desarrollo cognitivo individual (ontogénesis), transmitiendo conocimientos y habilidades (Apud, 2014). Así, surge una dialéctica entre la mente humana y las tecnologías, donde el uso de estas últimas implica transformaciones cognitivas, ya que, aunque las tecnologías actuales no alteran nuestra biología, sí alteran la plasticidad de la mente (Apud, 2014).

En la actualidad, el desarrollo de las TICs ha acelerado la aparición de herramientas que amplifican nuestra capacidad cognitiva, facilitando el acceso inmediato a información, la comunicación a distancia y la automatización de ciertos procesos sin necesidad de dominar conocimientos específicos (Apud, 2014). Este cambio suscita nuevas preguntas sobre la relación entre cognición y tecnología y hasta qué punto una puede transformar a la otra.

Además, plantea el tema de la distribución de funciones cognitivas en artefactos externos, lo que nos invita a examinar cómo éstos impactan en las capacidades mentales de los usuarios y cómo se ven modificadas por su uso.

La Teoría de la Mente Extendida (TME), desarrollada por Andy Clark y David Chalmers (1998), se enmarca dentro de la llamada nueva ciencia cognitiva. Esta perspectiva, que también integra influencias de filósofos como Husserl, Heidegger y Merleau-Ponty o Hegel (Broncano, 2006) se distingue de la ciencia cognitiva clásica en que entiende la mente no sólo como un sistema interno similar a un ordenador, sino como un proceso enactivo, embebido, encarnado y extendido en el entorno[1]. En esta visión, la cognición está interconectada con el entorno y el cuerpo de manera que ciertos procesos externos —cuando están adecuadamente acoplados— se consideran parte del sistema cognitivo (Pérez, 2010). Así, la mente se extiende más allá de los límites físicos del cerebro, incorporando elementos externos como parte de su funcionamiento.

La relevancia de esta teoría en el contexto actual se vuelve evidente al observar cómo los dispositivos digitales, desde teléfonos inteligentes hasta interfaces cerebro-computadora, actúan como "vehículo de contenido"[2] que no solo almacenan información sino que amplían las capacidades de procesamiento y organización mental. Por ejemplo, los teléfonos inteligentes se han convertido en elementos indispensables que operan como una "memoria externa" que facilita el acceso inmediato a la información, modificando así la forma en que las personas gestionan sus recuerdos y planifican sus actividades. En este proceso de externalización de funciones mentales, los dispositivos digitales cumplen una función similar a una operación mental interna por lo que se convierte en una extensión de la mente, en una "prótesis cognitiva"[3].

Si la TME está en lo cierto, la mente no se limita a procesar información internamente, sino que utiliza el entorno como un recurso para extender sus capacidades cognitivas. Broncano

[1] La nueva ciencia cognitiva rechaza la idea de que la mente se limite exclusivamente al interior del cerebro. En su lugar, propone que los estados y procesos mentales están "encarnados" (embodied), al incluir componentes corporales, y "embebidos" (embedded), pues el entorno próximo influye y hasta determina el funcionamiento mental. Además, son "enactivos" (enactive), ya que incluyen acciones del cuerpo, y "extendidos" (extended), al considerar ciertos procesos externos como parte de la cognición bajo condiciones adecuadas (Pérez, 2010).

[2] El término "vehículo de contenido", según Broncano (2006), es un elemento, ya sea interno o externo al cerebro, que porta y transmite información relevante para los procesos cognitivos. Es importante distinguir entre el vehículo, que es el soporte físico de la información, y el contenido mental, que es la información en sí misma, interpretada y comprendida por el sujeto.

[3] En el contexto de la mente extendida, las prótesis cognitivas cumplen una función semejante a la de las prótesis físicas, al sustituir o mejorar habilidades que de otro modo serían limitadas. Para Broncano (2006), la prótesis mental representa la integración de herramientas y recursos externos al sistema cognitivo, extendiendo así las capacidades de la mente.

(2006) describe este proceso como una "confianza informacional y cognitiva" (2006:112) en el medio, similar a la confianza práctica que depositamos en nuestro entorno para movernos y actuar. La descarga computacional es un proceso clave en esta relación. Broncano, basado en las ideas de Dennett (1996), sostiene que la mente humana, entendida como un sistema computacionalmente abierto, delega datos y procesos computacionales al entorno. Esto nos permite resolver problemas que, de otro modo, nuestras capacidades mentales limitadas no podrían abordar.

Siguiendo a Dennett (1996), la mente humana utiliza el entorno como un espacio de apoyo para procesar y almacenar información, trasladando la mayor cantidad de datos posible a elementos externos. La dependencia de estos recursos externos se intensifica a medida que externalizamos más información, lo que nos permite aumentar nuestras capacidades cognitivas al complementarlas con estas "prótesis" externas. De manera similar, Donald (1991) argumenta que los seres humanos están definidos por su capacidad para crear en el entorno un depósito de símbolos externos, los cuales facilitan y enriquecen la cognición (Broncano, 2006).

Así, con el uso constante de la herramienta, se "internaliza", volviéndose transparente y automática, integrándose al sistema cognitivo como un módulo interactivo con las redes neuronales (Broncano, 2006). Clark y Chalmers (2011) también enfatizan la importancia de una conexión fiable para que un recurso externo se considere parte del sistema cognitivo. Si una herramienta está siempre disponible cuando se necesita, se integra de forma fiable a la red de recursos cognitivos del individuo creando una "fusión" entre la mente y el dispositivo externo, formando una unidad mental (Broncano, 2006).

Para determinar qué dispositivos pueden considerarse parte del sistema cognitivo, Clark y Chalmers propusieron el "principio de paridad", que sugiere que, si un proceso externo es funcionalmente similar a un proceso mental interno, además de confiable, accesible y fluido, podría considerarse como parte de la mente extendida y formar parte del sistema cognitivo. Por ejemplo, un estudiante que utiliza un cuaderno digital para organizar sus notas y tareas puede llegar a confiar plenamente en este dispositivo para estructurar su memoria y planificación. Con el tiempo, este cuaderno no solo actúa como un repositorio de información, sino que se convierte en una extensión de su mente, permitiéndole delegar parte de sus funciones cognitivas, como el recuerdo y la organización de ideas.

Sin embargo, Menary (2010) amplía esta noción, argumentando que la simple paridad funcional no basta; lo esencial es una integración cognitiva profunda, que involucra varios aspectos clave: Primero, Menary destaca la normatividad compartida: las herramientas

externas deben cumplir roles establecidos culturalmente dentro de las prácticas cognitivas, como el uso de cuadernos, calendarios y dispositivos digitales, los cuales se integran a la cognición por responder a normas sociales sobre la organización de información, el recuerdo de eventos o el cálculo de datos. Luego, señala la mediación cultural en la integración de estas herramientas dado que las prácticas, hábitos y valores culturales moldean cómo las entendemos e incorporamos en nuestras actividades cognitivas. Finalmente, Menary subraya que esta integración requiere participación activa y sincronización entre el individuo y la herramienta, de modo que la herramienta no solo reemplace una función interna, sino que se adapte activamente al usuario y su contexto, transformando los procesos cognitivos.

Para profundizar en qué circunstancias un dispositivo puede considerarse una Tecnología de la Mente Extendida (TME), es fundamental aclarar el concepto de "mente" y diferenciarlo del "cerebro". Mientras que el cerebro es una entidad física, la mente se entiende como un concepto que abarca no solo los procesos cognitivos, sino también componentes afectivos, motivacionales y sociales que influyen en el pensamiento. Esta distinción es clave, ya que a pesar de que muchos diseños tecnológicos actuales están orientados principalmente a extender capacidades cognitivas, también podrían diseñarse explícitamente para ampliar otros procesos, como los afectivos (Reiner y Nagel, 2017).

Un ejemplo expuesto por Peter B. Reiner y Saskia K. Nagel (2017) habla de un dispositivo que actúa como soporte de la mente es el GPS. Tomemos el caso de John, quien comienza a trabajar como taxista en Londres sin tener un conocimiento profundo de la ciudad. Al principio, el GPS de su teléfono funciona como un apoyo para orientarse y navegar, pero con el tiempo, a medida que sigue utilizando esta herramienta sin fallos, el GPS se integra a su proceso mental. Ahora, cuando una dirección y una ruta aparecen en la pantalla, John las sigue sin cuestionarlas, lo cual indica que el GPS ha pasado de ser un recurso externo a una extensión mental, actuando como una TME en el funcionamiento de su mente.

No obstante, existe un elemento adicional en este proceso: el algoritmo que se convierte en parte de la mente de John también representa los intereses de la empresa que diseñó la aplicación GPS. No sería extraño que esta empresa recibiera ingresos por redirigir a John hacia ciertos negocios, lo cual introduce un objetivo secundario en su relación con el GPS. Este conflicto de intereses pone en relieve que los TME pueden llegar a violar la autonomía del usuario. Además, la influencia del algoritmo puede volverse más sutil y difícil de detectar, haciendo que las vulneraciones a la autonomía sean aún más complejas (Reiner y Nagel, 2017). La creciente integración de algoritmos en la vida diaria desafía las ideas tradicionales sobre identidad y autonomía, y parece inevitable que cada vez más personas

comiencen a ver sus dispositivos como TME, aunque no lo conceptualicen explícitamente de esa forma (Reiner y Nagel, 2017).

Como vemos, esta relación dinámica plantea complejos desafíos éticos relacionados con la autonomía, la privacidad y la dependencia tecnológica, ya que los dispositivos pueden influir en nuestra cognición y ser vulnerables a manipulaciones externas. Reconocer la mente como una entidad extendida no solo permite una comprensión ampliada de la mente humana en relación con el entorno, sino que también nos desafía a reconsiderar la responsabilidad moral y la autonomía en un mundo cada vez más interconectado con dispositivos y algoritmos.

2.2. Privacidad mental

La privacidad mental se convierte en una preocupación cuando consideramos que los dispositivos tecnológicos pueden formar parte de la mente extendida de un individuo. El filósofo Michael Lynch destaca que la "privacidad del pensamiento" radica en el acceso exclusivo del individuo a sus propios pensamientos, lo que fundamenta su autonomía (Carter y Palermos, 2016; Lynch, 2013). Imaginemos que María utiliza un gestor de contraseñas en la nube como una extensión de su memoria funcional. Si alguien accediera a este recurso sin su consentimiento, podría invadir su privacidad cognitiva y vulnerar su autonomía, exponiendo información que ella considera parte de su vida mental. Acceder a esta información en el contexto de la TME sería como "leer la mente" (Clowes et al., 2024).

Palermos (2023) sostiene que los datos almacenados en dispositivos externos, como extensiones de la mente, deben estar protegidos con el mismo rigor que las actividades mentales privadas del cerebro. Propone controles estrictos, tanto legales como tecnológicos, para garantizar que solo el propietario tenga acceso, fortaleciendo la ciberseguridad y revisando las leyes vigentes. Este enfoque redefine la privacidad en la era de la mente extendida, subrayando la necesidad de proteger la autonomía y prevenir accesos no autorizados a lo que metafísicamente forma parte de nuestra mente.

Técnicas como la anonimización de datos y la identificación biométrica, aunque útiles, son insuficientes en el contexto de la TME. Por ejemplo, una empresa que almacena datos anonimizados podría utilizar los algoritmos para predecir preferencias de consumo o políticas y, por tanto, influir en los comportamientos y decisiones de los usuarios, planteando dilemas éticos sobre su manipulación. Si estos datos forman parte de la mente extendida, alterarlos equivale a intervenir en las mentes de los usuarios, lo que cuestiona la ética del uso

indiscriminado de estos datos, especialmente cuando los usuarios no consienten plenamente su empleo para fines comerciales (Becker et al., 2023; Clowes et al., 2024).

Estas preocupaciones revelan una dimensión adicional: la manipulación mental. Imaginemos el caso de un usuario de un asistente de IA que organiza su calendario y recuerda citas importantes. Si alguien accediera a este sistema y modificara intencionadamente la información almacenada, podría alterar sus planes y decisiones, afectando directamente sus creencias y comportamiento (Clowes et al., 2024). Según Carter (2021), esta manipulación puede tomar la forma de adquisición, introduciendo nuevas creencias, o de erradicación, eliminando creencias existentes. Aunque en contextos cotidianos podría parecer inofensiva, cuando se realiza de manera encubierta plantea serios dilemas éticos. Así como rechazamos manipulaciones no consentidas en los procesos mentales internos, este principio debe aplicarse también a los recursos cognitivos externos para proteger la autonomía y evitar intervenciones sin consentimiento explícito.

Imaginemos un caso más complejo: un ciclista utiliza una aplicación de navegación que clasifica rutas como “seguras” o “peligrosas”, pero un error en el algoritmo empieza a marcar rutas riesgosas como seguras. Aunque el fabricante corrige el problema, no informa al usuario, lo que podría interpretarse como manipulación mental, ya que altera las creencias del ciclista, quien confía plenamente en las recomendaciones, sin su conocimiento. Informar al usuario sería ético para respetar su autonomía, pero implicaría acceder a sus datos de uso, lo que podría percibirse como una invasión de privacidad. Este dilema ilustra la tensión entre garantizar la seguridad, respetar la autonomía y proteger la privacidad mental (Clowes et al., 2024).

En última instancia, el contexto de la mente extendida exige un replanteamiento integral de la privacidad y la ética en el uso de tecnologías cognitivas. La protección de los dispositivos externos, como extensiones de la mente, requiere tanto medidas legales como soluciones tecnológicas que aseguren la autonomía de los usuarios frente a manipulaciones y accesos no autorizados. Además, es esencial que estas medidas no se limiten a proteger datos, sino que también consideren los posibles impactos sobre la integridad de las creencias y el bienestar del individuo.

2.3. Autonomía

La autonomía, entendida como la capacidad de tomar decisiones libres de influencias indebidas (Frankfurt, 1971; Dworkin, 1981), es un pilar fundamental de la autodeterminación. Sin embargo, la introducción de tecnologías cognitivas extendidas plantea nuevos retos para la autonomía. Mientras estas herramientas prometen fortalecer nuestra capacidad para actuar de manera eficiente, también pueden limitar nuestra libertad al influir en nuestras decisiones y creencias según patrones algorítmicos. Aunque la dependencia tecnológica puede ser útil, también corre el riesgo de limitar nuestra capacidad de autodesarrollo al confinarnos dentro de límites predefinidos, lo que plantea un desafío ético clave para preservar la autonomía en la era digital.

La privacidad mental, íntimamente ligada a la autonomía, no solo resguarda pensamientos privados, sino que también protege valores esenciales como la libertad de pensamiento y la individualidad psicológica. Según McCarthy-Jones (2019), la falta de privacidad mental puede generar autocensura y presiones de conformidad, erosionando nuestra capacidad para desarrollar ideas propias. Blitz (2010) refuerza esta idea al argumentar que la libertad de expresión, considerada una manifestación de la libertad de pensamiento, es crucial para preservar la autonomía mental. En este contexto, la privacidad mental adquiere un carácter ético y jurídico indispensable para salvaguardar la capacidad de las personas de decidir quiénes quieren ser en una sociedad altamente dependiente de tecnologías (Palermos, 2023).

Reiner y Nagel (2017) identifican tres factores que impactan la percepción de violación a la autonomía en este contexto: la persuasión algorítmica, la importancia de la decisión y la capacidad del algoritmo para personalizarse según el usuario. La persuasión puede variar desde niveles mínimos hasta un grado de influencia casi coercitiva, mientras que la seriedad de la decisión depende del impacto potencial para el individuo. Finalmente, si el algoritmo adapta su comportamiento al aprendizaje de las preferencias del usuario, se percibe una menor invasión a la autonomía, ya que refleja mejor los deseos y necesidades del usuario.

Imaginemos el caso de Raquel, usuaria de una aplicación de bienestar mental basada en IA. La aplicación accede a datos sensibles, como estados de ánimo y patrones de actividad, para personalizar recomendaciones y reflexiones diarias que influyen en las decisiones de los usuarios. En este contexto, la importancia de la privacidad mental es evidente. La exposición de los pensamientos y emociones privados de los usuarios podría someterlos a una “presión de conformidad”, por ejemplo, podría sugerirles continuamente pensamientos “positivos” o evitar temas delicados para “mejorar” su bienestar. Esta tendencia podría inducir autocensura

y limitar la autenticidad de sus reflexiones internas, erosionando su capacidad de autonomía mental.

Además, la aplicación podría guiar sutilmente a Raquel hacia una dependencia de sus recomendaciones. La influencia del algoritmo podría ser tan persuasiva que, en situaciones importantes —por ejemplo, ante una crisis emocional— el usuario podría tomar decisiones basadas principalmente en las sugerencias de la aplicación. Esto compromete tanto su individualidad psicológica como su capacidad para explorar libremente sus pensamientos, ya que se estaría ajustando a un molde específico promovido por la aplicación.

En este contexto, Vold y Hernández-Orallo (2022) analizan los llamados “extensores de IA”. Estos dispositivos, capaces de anticipar necesidades mediante el rastreo de comportamientos optimizan el acceso a información relevante. Sin embargo, esta automatización puede tener efectos adversos. Imaginemos a Luis, quien utiliza un extensor de IA para mejorar sus hábitos personales. Si el extensor prioriza información basada en hábitos pasados en lugar de apoyar nuevos objetivos, como cambiar su dieta, puede sabotear su capacidad de autotransformación. Este tipo de intervención subraya cómo las tecnologías, aunque útiles, pueden obstaculizar la capacidad de redefinirnos, un aspecto central de la autonomía.

Este problema no se limita solamente al ámbito de los extensores de IA; también se presenta en una variedad de tecnologías “inteligentes”, especialmente aquellas que perfilan usuarios y personalizan contenido y están gobernadas por el funcionamiento de algoritmos. Delacroix y Veale (2020) advierten que estas herramientas pueden moldear la identidad de los usuarios según sus propias interpretaciones, fomentando “profecías autocumplidas” y limitando la autodeterminación (Clowes et al 2024:11). En lugar de facilitar la autoexploración, estas tecnologías podrían imponer una narrativa que refuerce lo que "somos" según sus análisis, afectando nuestra capacidad de decidir qué queremos ser.

En este sentido, Lavazza (2021) advierte sobre el riesgo de abusos por parte de instituciones o empresas que podrían intervenir en la cognición extendida para imponer tratamientos o influencias externas. Esto subraya la necesidad de establecer límites claros que protejan tanto la autonomía como la privacidad de los individuos, asegurando que puedan decidir libremente qué tecnologías integrar en su vida y cómo interactuar con ellas.

En conclusión, la autonomía está profundamente vinculada a la privacidad mental, que asegura nuestra capacidad para pensar libremente y mantener nuestra individualidad. Como hemos visto, la falta de privacidad puede provocar autocensura y reducir la libertad de pensamiento (McCarthy-Jones, 2019), mientras que la libertad de expresión refuerza la

autonomía mental (Blitz, 2010), algo esencial para una libertad de pensamiento (Palermos, 2023). La clave radica en equilibrar el diseño y uso de estas herramientas para que, en lugar de limitar nuestra libertad, refuercen nuestra capacidad de decidir quiénes somos y quiénes queremos llegar a ser.

2.4. Subjetividad y autoridad epistémica

La subjetividad y la autoridad epistémica constituyen pilares fundamentales para comprender cómo los seres humanos mantienen control sobre sus procesos cognitivos en un entorno donde las herramientas externas desempeñan un papel cada vez más relevante. Mientras que la subjetividad se vincula a la identidad personal y al control de los propios estados mentales, la autoridad epistémica se define como la capacidad del sujeto para supervisar, evaluar y validar sus conocimientos y creencias. Ambas nociones están intrínsecamente ligadas a la autonomía, entendida como la facultad de actuar libre y responsablemente.

En el marco de la TME, estas ideas adquieren una nueva dimensión. Tradicionalmente, la subjetividad se consideraba un fenómeno interno, confinado al cerebro. Sin embargo, la TME plantea que la subjetividad se amplía hacia el entorno cuando el individuo integra dispositivos externos en su cognición (Broncano, 2006; Pérez, 2010). Broncano (2006) redefine la subjetividad como un sistema funcional que incorpora elementos externos en la cognición, siempre que el individuo mantenga agencia y responsabilidad.

Así, el Yo no se limita a procesos internos del cerebro; más bien, se extiende al entorno mediante las "prótesis mentales" que se convierten en componentes constitutivos del sistema cognitivo. Esta perspectiva no solo amplía la noción de subjetividad, sino que también redefine la identidad personal como una construcción compartida en la interacción con tecnología y otros agentes. Pérez (2010) advierte que la expansión de la subjetividad hacia el entorno conlleva riesgos, como la dilución de los límites entre el sujeto y su entorno. Sin límites claros, la agencia y la identidad personal pueden verse comprometidas cuando las herramientas tecnológicas asumen funciones críticas sin garantizar el control del usuario.

La autoridad epistémica, definida como la capacidad del individuo para supervisar y validar sus propios estados mentales, se convierte en un eje central en este debate. Vega (2005) subraya que esta autoridad requiere que el sujeto mantenga un control autónomo sobre sus creencias y decisiones, incluso cuando estas se apoyan en dispositivos externos. Sin embargo, la opacidad o autonomía excesiva de estas herramientas puede debilitar la agencia humana, comprometiendo la responsabilidad del individuo y la privacidad mental.

Broncano (2006) establece tres condiciones para mantener la autoridad epistémica: en primer lugar, el individuo debe ser capaz de evaluarse como la fuente de autoridad sobre sus procesos cognitivos; en segundo lugar, la información debe integrarse en un marco coherente con las creencias y valores del sujeto; y, en tercer lugar, el sujeto debe conservar responsabilidad sobre sus acciones y decisiones. El incumplimiento de estas condiciones no solo afecta la autonomía del individuo, sino que también pone en riesgo la privacidad mental, un componente esencial de la identidad personal.

Imaginemos a Daniel, quien utiliza una aplicación de IA para tomar decisiones financieras. Esta herramienta analiza sus hábitos de gasto y sugiere inversiones personalizadas. Con el tiempo, Daniel comienza a confiar plenamente en las recomendaciones del sistema, sin cuestionarlas ni entender cómo se generan. Finalmente, al delegar completamente su responsabilidad en la herramienta, Daniel compromete su autonomía, quedando expuesto a posibles manipulaciones o errores del sistema. Este ejemplo ilustra cómo el incumplimiento de las condiciones propuestas por Broncano (2006) puede afectar la autoridad epistémica.

Así mismo, Pérez (2010) y Vega (2005) coinciden en que la relación crítica y autónoma del individuo con las herramientas externas es clave para preservar la autoridad epistémica. Cuando los sistemas tecnológicos se vuelven opacos o excesivamente autónomos, se corre el riesgo de delegar funciones cognitivas fundamentales a dispositivos que no poseen intencionalidad ni voluntad. En este sentido, la supervisión activa y el control sobre los sistemas externos son esenciales para garantizar que estos recursos funcionan como extensiones cognitivas y no como sustitutos de la agencia humana.

Por otro lado, si reconocemos que la identidad personal trasciende los límites corporales, debemos afrontar las implicaciones de intervenir en el entorno de una persona como si estuviéramos actuando directamente sobre su mente. La subjetividad y la identidad no son atributos aislados, sino construcciones dinámicas que emergen de sistemas intersubjetivos donde convergen relaciones sociales, compromisos éticos e interacciones tecnológicas (Broncano, 2006; Rovane, 1999). Según Broncano, la identidad personal no se limita a la coherencia interna del Yo, sino que depende de un entramado compartido con otros, en el que lo social y lo tecnológico desempeñan un papel crucial en la configuración del sujeto. Este enfoque redefine al individuo no como una entidad autónoma, sino como una red de conexiones que refleja su interacción con el mundo.

En esta línea, Rovane (1999) argumenta que la identidad personal es, en esencia, un concepto ético construido a partir de proyectos racionales que configuran una narrativa única para cada individuo. Tanto Rovane como Broncano coinciden en que la subjetividad y la identidad

personal emergen y se consolidan en un espacio público de reconocimiento mutuo. Este proceso se nutre de la interacción constante con otros agentes y con los elementos del entorno, subrayando la importancia de los compromisos éticos y la responsabilidad compartida en la definición de la identidad.

De este modo, la subjetividad y la identidad personal se configuran como sistemas dinámicos, sostenidos por relaciones interactivas y capacidades compartidas. La autoconciencia no es un proceso aislado; está imbricada en un marco de representación pública y afirmación social (Broncano, 2006). En este contexto, la percepción del cuerpo y la privacidad de la mente se reinterpretan como derechos inherentes a la identidad personal, trascendiendo lo fisiológico para abarcar dimensiones éticas, sociales y tecnológicas. Estos derechos incluyen la intimidad, el control del espacio mental y la capacidad de decidir qué elementos externos forman parte del Yo.

En conclusión, la TME redefine la subjetividad y la identidad personal en un marco interdependiente donde herramientas externas se integran como componentes activos de la cognición. Esto exige asumir una responsabilidad ética frente a las implicaciones de esta interdependencia, plantear la necesidad de establecer límites claros que preserven la autonomía y la agencia del individuo, evitando que la dependencia tecnológica comprometa su integridad. La responsabilidad ética radica en garantizar que las decisiones sobre la integración tecnológica promuevan el control crítico y el respeto por la privacidad.

3. Hacia una ética relacional de la responsabilidad

3.1. Dinámicas relacionales y retos éticos

La exploración realizada en el capítulo anterior ha evidenciado cómo las tecnologías cognitivas transforman dimensiones esenciales de la experiencia humana, como la privacidad mental, la autonomía y la subjetividad, al tiempo que reconfiguran las dinámicas de autoridad epistémica. Asimismo, se ha destacado que estas relaciones no ocurren de manera aislada, sino en un marco de interdependencia compleja donde la tecnología actúa como un agente activo en la configuración de las capacidades cognitivas, sociales y éticas. Este capítulo profundiza en las dinámicas relacionales de esta interacción humano-tecnológica y plantea los retos para un marco ético adaptado a estas nuevas realidades.

En este contexto, David Kirsh (2006) conceptualiza la interacción humano-tecnológica como un proceso de coevolución, donde las innovaciones tecnológicas no solo responden a las necesidades humanas, sino que también las redefinen y moldean nuestras capacidades. Esta visión se complementa con la teoría de la cognición distribuida de Edwin Hutchins (2000), que expone cómo la mente humana opera en redes de interacción entre personas, herramientas y entornos. Estas ideas refuerzan el análisis previo al señalar que las tecnologías cognitivas actúan como extensiones de la mente humana, redistribuyendo las tareas cognitivas y alterando la forma en que construimos conocimiento y tomamos decisiones.

De manera interrelacionada, este fenómeno puede entenderse también desde la perspectiva de la ecología cognitiva, desarrollada por autores como Apud (2014). Este enfoque destaca que la cultura y el entorno tecnológico no son simplemente colecciones de objetos, sino sistemas dinámicos que estructuran y distribuyen el conocimiento de manera activa. Al igual que la cognición distribuida, la ecología cognitiva subraya cómo las interacciones constantes entre la mente humana y el entorno influyen directamente en el desarrollo de nuestra cognición.

En este marco de interdependencia, las herramientas tecnológicas dejan de ser elementos neutrales para convertirse en mediadores que incorporan intenciones y valores humanos, configurando nuestra percepción, acción y toma de decisiones (Llinares, 2018). Esta visión resalta la necesidad de que la ética de la mente extendida aborde críticamente el diseño y propósito de estas herramientas, entendiendo la tecnología como un medio activo que cristaliza las intenciones de sus creadores e impacta en la cognición y el comportamiento de los usuarios (Apud, 2014).

Construir un marco ético, por tanto, requiere comprender los contextos en los que humanos y tecnología coevolucionan, identificando cómo estas interacciones configuran las prácticas morales y las responsabilidades compartidas. Es en estos espacios concretos —donde la tecnología se utiliza, adapta y se convierte en una parte integral de nuestras acciones— donde se forjan las capacidades éticas y el sentido de responsabilidad hacia los demás y el entorno.

La autonomía relacional, propuesta por Jennifer Nedelsky (2011) y Yael Braudo-Bahat (2017), se presenta como un concepto clave en este marco de interdependencia humano-tecnológica. Este enfoque redefine la autonomía no como independencia absoluta, sino como una capacidad que se construye y se sostiene en interacción con otros agentes y con el entorno. Nedelsky destaca que las relaciones constructivas son aquellas que fomentan el pensamiento crítico, la creatividad y la autodeterminación, mientras que las relaciones destructivas restringen estas capacidades mediante coerción o dependencia excesiva.

Además, esta perspectiva destaca que la autonomía no solo depende de las decisiones racionales de un individuo, sino también de su capacidad para involucrarse en relaciones que proporcionen apoyo emocional, social y ético. Según Braudo-Bahat (2017), la autonomía relacional no niega la individualidad, sino que la integra en una red de interdependencia, enfatizando cómo las decisiones y valores personales se enriquecen a través del diálogo y el aprendizaje mutuo dentro de estas relaciones.

Desde esta perspectiva, las tecnologías cognitivas deben diseñarse para ser aliadas en la construcción de autonomía relacional. No obstante, existen desafíos significativos. Cuando las tecnologías operan de manera opaca o imponen dinámicas de dependencia acrítica, debilitan la capacidad de los usuarios para actuar de manera reflexiva e informada. Este es un riesgo ético central, ya que las tecnologías cognitivas pueden fortalecer o socavar la autonomía dependiendo de su diseño y uso.

Por ejemplo, un sistema de IA que asista en decisiones médicas puede empoderar a los profesionales al proporcionar análisis de datos avanzados. Sin embargo, si este sistema opera de manera opaca y fomenta la dependencia, puede debilitar la capacidad crítica del usuario, convirtiéndolo en un mero ejecutor de decisiones algorítmicas. Este dilema subraya la necesidad de diseñar tecnologías que respeten y refuercen la agencia del usuario, garantizando que puedan supervisar, cuestionar y validar las decisiones tecnológicas.

Por tanto, el reto ético que plantea este capítulo no es simplemente evaluar el impacto de las tecnologías cognitivas, sino construir un marco ético que sea dinámico, adaptable y coherente con las características relacionales de la interdependencia humano-tecnológica. Este marco debe estar fundamentado en la autonomía relacional como pilar central, reconociendo que las decisiones éticas no pueden desvincularse de las relaciones en las que están inscritas. Es en estas interacciones donde se configuran las prácticas morales y los valores que deberían orientar tanto el diseño como el uso de tecnologías cognitivas.

Además, un marco ético basado en la autonomía relacional debe ser lo suficientemente flexible para responder a los cambios rápidos y disruptivos que caracterizan la innovación tecnológica. Esto incluye la capacidad de anticipar riesgos, garantizar la equidad en la distribución de beneficios y responsabilidades, y preservar las capacidades críticas y reflexivas de los usuarios. Estas características serán analizadas críticamente en el próximo capítulo, donde se evaluarán los marcos éticos tradicionales y sus limitaciones para abordar las complejidades de la tecnología contemporánea.

3.2. Ética relacional de la responsabilidad

La creciente complejidad de la relación entre humanos y tecnologías cognitivas, analizada en el capítulo anterior, pone de manifiesto la necesidad de un cambio de paradigma ético que responda a las dinámicas de interdependencia y coevolución que caracterizan estas interacciones. Aunque los marcos éticos tradicionales —como el consecuencialismo, la deontología y la ética de la virtud— han sido fundamentales para abordar dilemas éticos en contextos humanos, presentan limitaciones significativas frente a los retos de una tecnología que actúa como extensión de la mente y mediadora en nuestras capacidades cognitivas, sociales y éticas.

El consecuencialismo, especialmente en su versión utilitarista, evalúa las acciones según sus resultados, priorizando aquellas que maximizan el bienestar colectivo (Mill, 1863; Sinnott-Armstrong, 2021). Este enfoque puede ser útil en contextos donde los efectos de una tecnología son medibles y cuantificables, como en aplicaciones destinadas a mejorar la productividad o reducir costos. Sin embargo, su énfasis en resultados agregados tiende a ignorar aspectos cualitativos críticos, como la autonomía individual, la privacidad mental o las relaciones de poder que pueden emerger en redes tecnológicas.

Por ejemplo, un algoritmo de recomendación que optimiza el tiempo de visualización en plataformas de contenido podría considerarse exitoso desde un enfoque utilitarista, pero al mismo tiempo, este sistema puede socavar la autonomía del usuario al manipular sus preferencias de manera opaca y al reforzar sesgos cognitivos. Además, el carácter agregacionista del consecuencialismo puede justificar decisiones que sacrifican los intereses de minorías en favor de un supuesto bienestar general, lo cual resulta éticamente problemático en contextos donde las tecnologías perpetúan desigualdades estructurales o erosionan derechos fundamentales (Cortina, 2024).

La deontología, por su parte, se centra en el respeto a principios y deberes universales, como la autonomía, la dignidad y los derechos humanos (Kant, 1785; Alexander y Moore, 2021). Desde esta perspectiva, las tecnologías cognitivas deberían diseñarse y utilizarse de manera que respeten siempre a los individuos como fines en sí mismos y nunca como simples medios para alcanzar otros objetivos. Un ejemplo sería un sistema de inteligencia artificial que no manipule la toma de decisiones del usuario ni vulnere su privacidad mental.

Sin embargo, este enfoque también enfrenta críticas por su rigidez. La deontología a menudo carece de la flexibilidad necesaria para adaptarse a la evolución tecnológica, lo que dificulta anticipar cómo las innovaciones afectan los principios éticos (Vallor, 2016). Por ejemplo,

proteger la privacidad de los datos puede entrar en conflicto con la necesidad de recopilar información para optimizar el funcionamiento de un sistema. Además, la deontología carece de mecanismos prospectivos que permitan anticipar y mitigar los riesgos futuros derivados del desarrollo y uso de estas tecnologías. Esto es particularmente problemático en entornos donde las decisiones tecnológicas tienen consecuencias acumulativas e imprevisibles, como en sistemas de IA que evolucionan de manera autónoma.

Finalmente, la ética de la virtud, a diferencia de los enfoques mencionados, se centra en el desarrollo del carácter moral y en el florecimiento humano, evaluando las acciones según los valores y las virtudes que promueven, como la prudencia, la compasión y la justicia (Clowes et al., 2024). En el contexto de las tecnologías cognitivas, este enfoque podría guiar a diseñadores y usuarios hacia prácticas responsables que consideren el impacto de las herramientas en el bienestar integral de las personas y su dignidad. Por ejemplo, una empresa que desarrolla una inteligencia artificial educativa debería reflexionar sobre cómo su producto promueve habilidades críticas y autonomía en los estudiantes, en lugar de fomentar la dependencia o la pasividad cognitiva.

Sin embargo, la ética de la virtud enfrenta desafíos significativos en entornos tecnológicos. Su énfasis en valores abstractos no proporciona un marco sistemático para resolver conflictos éticos en sistemas distribuidos y opacos, donde los impactos de las decisiones no siempre son evidentes o inmediatos. Por ejemplo, el diseño de un sistema educativo basado en inteligencia artificial podría alinearse con el valor de promover el pensamiento crítico, pero ¿cómo garantizar que este objetivo se cumpla cuando las dinámicas de uso y los resultados del sistema dependen de múltiples actores y contextos?

Estas limitaciones ponen en evidencia que los marcos éticos tradicionales, aunque valiosos en sus propios contextos, no contemplan las necesidades de una ética adaptada a las complejidades de las tecnologías cognitivas contemporáneas. La interdependencia y distribución de funciones en redes humanas y tecnológicas requieren una ética que no solo sea dinámica y adaptativa a contextos particulares y rápidos avances tecnológicos, sino que también incorpore una visión prospectiva capaz de anticipar impactos futuros para garantizar el bienestar y la protección de todos los actores dentro de estas redes interdependientes.

Los enfoques tradicionales tienden a priorizar perspectivas imparciales e individualistas (Gilligan, 2013), lo que resulta insuficiente en un contexto donde las decisiones y sus consecuencias están distribuidas entre múltiples actores, tanto humanos como tecnológicos. Adela Cortina (2024) argumenta que una sociedad verdaderamente comprometida con la dignidad humana, el bienestar y la protección de las personas debe adoptar una ética basada

en la intersubjetividad. Frente a este escenario, se requiere una ética intersubjetiva que reconozca y gestione las responsabilidades compartidas, fomente la cooperación entre los actores y aborde las tensiones éticas inherentes a la mediación tecnológica.

En este contexto, surge la necesidad de una ética relacional de la responsabilidad, un marco que integre los elementos más valiosos de estos enfoques y los adapte a las particularidades de las tecnologías cognitivas en el marco de la teoría de la mente extendida. Esta ética reconoce que los seres humanos no son entidades aisladas; nuestra identidad y desarrollo personal surgen del reconocimiento mutuo y de las relaciones que establecemos con los demás (Cortina, 2024).

Desde esta perspectiva, entendemos que la responsabilidad no es una carga individual, sino una práctica relacional y distribuida que emerge de la interacción entre humanos, tecnologías y contextos sociales. Esta idea se alinea con la postura de Broncano (2006), quien sostiene que la subjetividad y la identidad humana no son atributos individuales, sino construcciones relacionales que emergen en interacción con otros agentes y con el entorno. Además, es importante que la responsabilidad sea prospectiva, asumiendo la necesidad de evaluar no solo las consecuencias inmediatas de nuestras decisiones, sino también los posibles riesgos y beneficios que puedan emerger en el largo plazo.

Además, la ética relacional de la responsabilidad se basa en la noción de la autonomía relacional que hemos visto en el anterior capítulo (Nedelsky, 2011; Braudo-Bahat, 2017). Lo que plantea que las tecnologías deben diseñarse para ampliar la capacidad crítica y reflexiva de los usuarios, fomentando decisiones informadas en redes socio-tecnológicas complejas. Según Braudo-Bahat (2017), esto requiere un entorno que facilite opciones diversas y permita decisiones auténticas, promoviendo el diálogo inclusivo y la construcción colectiva del conocimiento. Así, este enfoque conecta la ética relacional con el diseño de tecnologías que respeten la autonomía y favorezcan interacciones justas entre humanos y artefactos.

Así, la ética relacional de la responsabilidad se presenta como una respuesta integral, al reconocer la interdependencia y la distribución de funciones en redes socio-tecnológicas, donde las decisiones y sus consecuencias están compartidas entre múltiples actores. Este enfoque resalta la dimensión intersubjetiva de la identidad y la subjetividad, al mismo tiempo que pone de relieve una responsabilidad prospectiva, orientada a anticipar escenarios futuros.

La ética relacional enfatiza que las tecnologías deben diseñarse para ampliar la capacidad crítica y reflexiva de los usuarios, promoviendo decisiones informadas y auténticas en entornos complejos. Este planteamiento sienta las bases para los principios éticos del

siguiente capítulo, que profundizan en cómo garantizar claridad, equidad y justicia en el diseño y uso de tecnologías cognitivas, favoreciendo interacciones equilibradas y respetuosas entre humanos y artefactos.

3.3. Principios éticos

La exploración previa ha subrayado la creciente interdependencia entre humanos y tecnologías cognitivas, presentando riesgos como la erosión de la privacidad mental, la dependencia tecnológica, la manipulación algorítmica y la pérdida de autonomía. Además, hemos analizamos cómo estas dinámicas relacionales reconfiguran la subjetividad y la autoridad epistémica, posicionando a las tecnologías como mediadoras activas en nuestras capacidades cognitivas y sociales. Hemos comprobado que los marcos éticos tradicionales no han logrado abordar completamente estos retos. Por ello, establecimos la necesidad de una ética que reconozca la relacionalidad y responsabilidad en las redes socio-tecnológicas.

I. Principio de claridad y autonomía

El principio de claridad y autonomía surge como una respuesta ética fundamental para garantizar que las tecnologías cognitivas respeten y fortalezcan las capacidades críticas y reflexivas de los usuarios. A diferencia de la simple transparencia, que se limita a proporcionar acceso a información técnica (Parlamento Europeo, 2023), la claridad exige que esta información sea comprensible y accesible para los usuarios en función de su contexto y capacidades. Este principio reconoce que la autonomía no puede mantenerse sin herramientas adecuadas para interpretar las decisiones tecnológicas, ya que los usuarios se verían relegados a un rol pasivo frente a sistemas que influyen directamente en ellos.

Además, la claridad desempeña un papel esencial en la autoridad epistémica, es decir, en la capacidad del usuario para cuestionar y validar las decisiones algorítmicas. Según Broncano (2006), la autoridad epistémica requiere que los individuos se perciban como fuentes de control sobre su actividad mental, que la información sea coherente con sus valores y creencias, y que conserven la responsabilidad sobre sus acciones. Por ejemplo, si un sistema educativo basado en IA no explica los criterios detrás de sus recomendaciones, podría imponer modelos de conocimiento que no reflejen las necesidades o valores del individuo, debilitando su capacidad crítica y reflexiva.

La privacidad mental que, como se analizó anteriormente, abarca la salvaguarda de los procesos cognitivos internos que las tecnologías pueden mediar o influenciar (Palermos, 2023) también podría verse afectada por una falta de claridad. Por ejemplo, los sistemas de inteligencia artificial que predicen patrones de comportamiento, deben explicar claramente cómo recopilan y utilizan datos para evitar vulneraciones involuntarias de la privacidad mental. Una falta de claridad en estos procesos puede derivar en dinámicas de dependencia y manipulación algorítmica, comprometiendo la autonomía del usuario.

La explicabilidad, como dimensión esencial de la claridad, asegura que los usuarios puedan comprender y cuestionar los sistemas tecnológicos. Sin esta comprensión activa, existe el riesgo de que las tecnologías asuman un papel epistémico dominante, comprometiendo la autonomía relacional de los usuarios. Esto también es un antídoto frente a la opacidad técnica inherente a los sistemas complejos, como los algoritmos de aprendizaje profundo, que funcionan como una "caja negra", en la que incluso los desarrolladores pueden desconocer los procesos internos que generan resultados (Burrell, 2016).

Por otro lado, aunque la transparencia busca ofrecer información, esta puede resultar demasiado técnica o extensa, lo que puede provocar una sobrecarga cognitiva y dificultar la toma de decisiones informada. El AI Act Europeo[4] aborda este problema al establecer que las explicaciones proporcionadas por los sistemas de IA deben ser claras, accesibles y adaptadas a las necesidades específicas de cada grupo (Parlamento Europeo y Consejo de la Unión Europea, 2024). Por ejemplo, mientras los usuarios finales requieren explicaciones simplificadas que les permitan comprender cómo una IA toma decisiones que afectan su vida cotidiana, los técnicos necesitan descripciones más detalladas sobre los algoritmos y el manejo de datos. Por su parte, los reguladores demandan información estructurada que permita auditar el sistema y garantizar su cumplimiento con la normativa vigente.

La claridad también se conecta con la dimensión ética de la confianza mutua entre los usuarios y las tecnologías. Las explicaciones claras y relevantes fortalecen esta confianza al garantizar que los usuarios comprendan y puedan prever cómo los sistemas tecnológicos afectan sus decisiones. Por ejemplo, en plataformas de redes sociales, una explicación clara de los algoritmos que priorizan contenido podría ayudar a los usuarios a gestionar mejor su

[4] El AI Act Europeo (Reglamento (UE) 2024/1689) es un marco legal adoptado por la Unión Europea para regular el diseño, desarrollo, comercialización y uso de sistemas de inteligencia artificial. Su objetivo es garantizar que estos sistemas sean seguros, transparentes y respeten los derechos fundamentales y valores éticos de la UE. Introduce un enfoque basado en el riesgo, clasificando los sistemas de IA en diferentes categorías según su impacto potencial, y establece requisitos específicos para aquellos considerados de alto riesgo. (Parlamento Europeo y Consejo de la Unión Europea, 2024).

exposición a información sesgada, fortaleciendo su capacidad crítica y evitando manipulaciones comerciales o políticas.

En conclusión, el principio de claridad y autonomía redefine la relación entre los usuarios y las tecnologías cognitivas, enfatizando la necesidad de proporcionar información comprensible, adaptada y relevante. Este enfoque no solo protege la privacidad mental y fortalece la autoridad epistémica, sino que también garantiza que los usuarios mantengan su capacidad crítica y reflexiva frente a decisiones tecnológicas complejas. Al integrar este principio en el diseño y uso de las tecnologías cognitivas, se promueve una interacción equilibrada y responsable que fomenta una autonomía relacional crítica y reflexiva.

II. Principio de responsabilidad interdependiente y justicia

El principio de responsabilidad interdependiente y justicia reconoce que en las redes complejas y distribuidas que caracterizan a las tecnologías cognitivas, la responsabilidad ética no puede reducirse a un único agente humano o institucional. Este principio subraya la necesidad de adoptar una perspectiva relacional que considere la agencia compartida entre diseñadores, usuarios, instituciones y las propias tecnologías. Además, destaca la importancia de garantizar que los beneficios y riesgos de estas tecnologías se distribuyan de manera justa, evitando la perpetuación de desigualdades estructurales.

La teoría de la agencia híbrida de Gunkel (2017) y la teoría del actor-red de Latour (Argemí y Tirado, s.f.) justifican esta perspectiva al redefinir la agencia como el producto de interacciones complejas en redes interdependientes de humanos y no humanos donde las responsabilidades no pueden atribuirse exclusivamente a humanos. Por ejemplo, un algoritmo de selección de personal no solo depende del programador que lo diseñó, sino también de los datos con los que fue entrenado, las políticas de la empresa y las decisiones del reclutador. Este entramado hace necesaria una responsabilidad distribuida en todas las etapas de su desarrollo y uso de la tecnología.

De manera complementaria, la ética híbrida de Verbeek (2011) plantea que la interacción humano-tecnológica debe entenderse como un proceso de co-creación en el que las tecnologías no solo amplían capacidades, sino que transforman valores, relaciones y decisiones. Este enfoque resalta que plataformas como las redes sociales, que influyen en la percepción y el comportamiento de los usuarios, generan responsabilidades compartidas entre desarrolladores, empresas y usuarios. Ignorar esta interdependencia fomenta dinámicas de poder opacas que erosionan la autonomía y perpetúan relaciones desiguales.

Desde esta perspectiva, Hanson (2009) introduce el concepto de agencia extendida, que desafía el individualismo moral y propone una visión ecológica de la ética, donde las tecnologías actúan como extensiones activas de la acción humana. Este marco refuerza la necesidad de gestionar éticamente las redes híbridas para evaluar constantemente las dinámicas de poder y control. Así, la responsabilidad interdependiente no es solo un reconocimiento teórico, sino una exigencia práctica para garantizar una interacción justa, transparente y equilibrada entre humanos y tecnologías.

La justicia es un eje central de este principio, ya que busca garantizar que las tecnologías no perpetúen desigualdades estructurales, sino que contribuyan a la equidad y la inclusión. Según Floridi (2024), la justicia implica corregir sesgos algorítmicos, garantizar accesibilidad y distribuir los beneficios de manera equitativa. Por ejemplo, los sistemas de IA utilizados en la justicia penal deben diseñarse y evaluarse para garantizar que no reproduzcan prejuicios raciales o de género en sus recomendaciones. La justicia tecnológica no solo implica acceso equitativo a las herramientas, sino también un análisis crítico de cómo estas impactan en el desarrollo cognitivo, la autodeterminación y la dignidad de los usuarios.

Uno de los mayores desafíos del principio de responsabilidad interdependiente es garantizar que las redes híbridas sean evaluadas de manera continua para evitar dinámicas de dominación tecnológica. Esto requiere la implementación de mecanismos de supervisión ética continua y auditorías externas que permitan identificar y mitigar riesgos emergentes. Además, las evaluaciones dinámicas deben incluir la participación activa de todos los actores involucrados, asegurando que las tecnologías permanezcan alineadas con los valores éticos y las necesidades de los usuarios.

Sin embargo, existe un desafío en este marco: la difusión de la responsabilidad. Esto puede dificultar la rendición de cuentas, especialmente en sistemas tecnológicos autónomos. Por ejemplo, cuando un sistema algorítmico toma decisiones en un ámbito como la justicia penal o la selección de personal, los usuarios finales, los desarrolladores y los diseñadores comparten responsabilidades que no siempre están claramente definidas. Esta ambigüedad puede contrarrestar la efectividad o dar excesiva autoridad a las tecnologías.

Resolver estas tensiones sigue siendo una cuestión crítica. En este contexto, Gunkel (2017) analiza la posibilidad de programar valores éticos en sistemas autónomos para que tomen decisiones moralmente informadas. Sin embargo, advierte que las máquinas carecen de una comprensión completa de los valores humanos, limitando la efectividad de la llamada “ética funcional”. Esta crítica recalca el papel esencial de los diseñadores y tecnólogos, no solo en

la programación de valores éticos, sino también en la integración de marcos éticos que guíen a todos los agentes involucrados.

Además, los conflictos de intereses entre los diversos actores pueden obstaculizar la implementación de medidas éticas justas y equitativas. Por ejemplo, las prioridades comerciales, como el beneficio económico, pueden entrar en conflicto con los valores éticos fundamentales, como la equidad y la justicia social. En este contexto, se requiere un marco ético que no solo reconozca la complejidad de estas redes híbridas, sino que también promueva mecanismos claros de asignación y rendición de cuentas.

Para abordar este desafío, una posible solución inspirada en el AI Act Europeo es la implementación de auditorías éticas obligatorias para evaluar el cumplimiento de las normativas y principios en el desarrollo y uso de tecnologías cognitivas. Estas auditorías deberían ser realizadas por entidades independientes y abarcar tanto los riesgos inmediatos como las repercusiones a largo plazo en los derechos fundamentales. Asimismo, el establecimiento de responsables designados para cada etapa del ciclo de vida de la tecnología, desde su diseño hasta su implementación y supervisión, puede garantizar que las responsabilidades no se diluyen entre los múltiples actores.

En conclusión, este principio redefine la relación ética entre humanos y tecnologías como un proceso relacional, colectivo y dinámico. Al integrar teorías como la agencia híbrida, el actor-red y la ética híbrida, este principio reconoce que la responsabilidad ética no es exclusiva de los humanos, sino que está distribuida en redes complejas de interacción. Además, subraya la importancia de garantizar que las tecnologías respeten la autonomía y la dignidad humanas, promoviendo la equidad y la justicia en su desarrollo y uso.

Para que este enfoque sea efectivo, es esencial adoptar prácticas de supervisión ética continua, regulación normativa clara y mecanismos de auditoría que aseguren una distribución equitativa de beneficios y riesgos. Solo así será posible construir un marco ético que responda a los desafíos planteados por las tecnologías cognitivas.

III. Principio de prospectiva ética y beneficencia

El principio de prospectiva ética y beneficencia se erige como una respuesta a los desafíos de anticipar y gestionar los riesgos éticos y sociales inherentes al desarrollo de tecnologías cognitivas. Inspirado en la filosofía de Hans Jonas (2008), este principio subraya la necesidad de asumir una responsabilidad prospectiva que trascienda el presente y contemple los efectos a largo plazo de las decisiones tecnológicas. Esta perspectiva se complementa con los pilares

éticos de beneficencia y no maleficencia, fundamentales para garantizar que las tecnologías no sólo eviten causar daño, sino que también contribuyan al bienestar humano y ambiental.

En este contexto, Jonas redefine el concepto kantiano de deber. Mientras que Kant proponía que "puedes, puesto que debes," Jonas lo adapta al "debes, puesto que haces, puesto que puedes" (Jonas, 2008, p. 212). Este principio subraya que con el aumento del poder humano crece también la responsabilidad proporcional de garantizar no solo la preservación de la existencia, sino también las mejores condiciones de vida para las generaciones futuras. El uso irresponsable de este poder compromete no solo el bienestar inmediato, sino también la dignidad humana y la sostenibilidad a largo plazo.

La anticipación de riesgos éticos y sociales se enfrenta al desafío de la imprevisibilidad tecnológica (Llinares, 2018). Como señala Jonas (2008), el poder transformador de la tecnología amplifica el alcance de nuestras acciones, exigiendo una reformulación ética que priorice la prevención de escenarios adversos (Yáñez, 2021). Jonas introduce la "heurística del temor" como un instrumento ético que permite imaginar posibles escenarios destructivos, no con el fin de paralizar la innovación, sino para fomentar una reflexión ética profunda que permita imaginar y mitigar posibles consecuencias negativas antes de que se materialicen. La anticipación del futuro se convierte, así, en una condición para la acción ética (Yáñez, 2021).

Este enfoque es particularmente relevante en contextos donde la naturaleza multiestable de la tecnología, como describe Albrechtslund (2007), dificulta prever todos sus posibles usos y efectos. La multiestabilidad de las tecnologías, que permite que adquieran múltiples significados y aplicaciones dependiendo del contexto, exige un diseño flexible y adaptativo. Por ejemplo, los sistemas de inteligencia artificial diseñados inicialmente para optimizar procesos administrativos pueden terminar utilizándose en contextos sensibles, como la justicia penal, donde sus implicaciones éticas son mucho más profundas y complejas.

Así mismo, el AI Act Europeo propone abordar la mitigación de riesgos mediante un enfoque basado en la identificación y reducción proactiva de los impactos negativos. Establece que los responsables del despliegue deben implementar medidas específicas como evaluaciones de impacto *ex ante*, diseño robusto y pruebas rigurosas en condiciones controladas antes de la implementación en entornos reales. Además, el Reglamento enfatiza la importancia de adaptar los sistemas a cambios contextuales mediante actualizaciones regulares y auditorías continuas, asegurando que se mantengan alineados con los valores y derechos fundamentales, incluso frente a riesgos emergentes e imprevisibles (Parlamento Europeo y Consejo de la Unión Europea, 2024). Esto conecta con las prácticas de falsabilidad y despliegue

incremental al garantizar la fiabilidad de la tecnología mediante pruebas empíricas y despliegues graduales en contextos seguros (Floridi, 2024).

La beneficencia, entendida como la promoción activa del bienestar humano y social, va más allá de la simple prevención del daño. Según Floridi (2024), este principio implica diseñar tecnologías que no solo respeten la dignidad humana, sino que también fortalezcan valores fundamentales como la privacidad mental, la equidad y la sostenibilidad. Por su parte, el principio de no maleficencia exige que los diseñadores, desarrolladores y usuarios se esfuercen por minimizar los riesgos y evitar usos indebidos que comprometan la seguridad, la autonomía o el bienestar cognitivo de los individuos.

En este contexto, el diseño participativo emerge como una herramienta fundamental para integrar los principios de beneficencia y no maleficencia en el desarrollo tecnológico. Este enfoque, alineado con las teorías de la agencia distribuida y el Diseño Sensible a los Valores (Friedman et al., 2002), implica la colaboración activa de múltiples actores durante todas las etapas del proceso de diseño. Al incorporar perspectivas diversas, garantiza que las tecnologías respondan a necesidades reales y equilibren valores éticos en potencial conflicto, como la privacidad y la claridad. Según Friedman et al. (2002), la integración de principios éticos en el diseño requiere un diálogo constante con todos los actores implicados, fortaleciendo así la capacidad de las tecnologías para reflejar un equilibrio adecuado entre valores sociales y capacidades tecnológicas.

Así mismo, es importante integrar los contextos históricos y locales en el diseño, tal y como señala Apud (2014). Este enfoque destaca que el diseño ético no puede ser universal ni homogéneo, sino que debe adaptarse a las particularidades culturales, sociales y económicas de las comunidades donde se implementan las tecnologías. Al considerar estos contextos, el diseño participativo no solo garantiza que las tecnologías sean funcionales, sino que también eviten dinámicas de exclusión o dominación tecnológica.

En este sentido, Gilligan (2013) argumenta que el cuidado, con su énfasis en la empatía y el reconocimiento de la diversidad, puede ofrecer una base para diseñar tecnologías que promuevan relaciones humanas respetuosas y equitativas. Integrar el cuidado como un valor central en el diseño tecnológico no sólo implica evitar el daño, sino también construir entornos que fortalezcan la agencia, el bienestar emocional y las relaciones interpersonales. Gilligan subraya, además, la importancia del reconocimiento de la voz del Otro como elementos clave para contrarrestar las jerarquías que deshumanizan las relaciones. Este enfoque promueve sistemas diseñados para ser inclusivos y sensibles a las diversas voces y contextos de los usuarios, asegurando que ninguna perspectiva quede marginada.

Imaginemos el diseño de una plataforma de aprendizaje en línea para comunidades rurales con acceso limitado a la tecnología. Desde un enfoque participativo, se involucraría a docentes, estudiantes, líderes comunitarios y expertos tecnológicos para identificar necesidades específicas, como la disponibilidad de contenido en lenguas indígenas y la accesibilidad en dispositivos de bajo costo. Incorporando los contextos locales y culturales, como señala Apud (2014), el diseño priorizaría valores comunitarios y herramientas adaptadas al entorno.

El principio de prospectiva ética y beneficencia se posiciona como una guía fundamental para anticipar los riesgos éticos y sociales de las tecnologías cognitivas, al mismo tiempo que promueve valores positivos que trascienden el presente. Al integrar la responsabilidad prospectiva de Jonas, los pilares de beneficencia y no maleficencia de Floridi, y los enfoques relacionales de Gilligan y otros autores, este principio ofrece un marco robusto para construir un futuro tecnológico más justo, inclusivo y sostenible. Sin embargo, su efectividad depende de la implementación de procesos dinámicos y participativos que garanticen que las tecnologías no solo eviten el daño, sino que también contribuyan de manera positiva al bienestar humano y social.

4. Conclusión

El presente trabajo ha tenido como objetivo principal evaluar los desafíos éticos derivados de la TME y proponer una ética relacional basada en la responsabilidad para guiar el diseño y uso de tecnologías cognitivas. A partir de este objetivo general, se plantearon cuatro objetivos específicos que han orientado el desarrollo de este trabajo: analizar los fundamentos teóricos de la TME, examinar el impacto de las tecnologías desde una perspectiva teórica y crítica, evaluar las limitaciones de los marcos éticos tradicionales y proponer un nuevo enfoque ético que responda a la complejidad humano-tecnológica. En este capítulo, se reflexiona sobre los logros y las limitaciones con dichos objetivos y se plantean posibles líneas futuras.

En primer lugar, el análisis de los fundamentos teóricos de la TME ha permitido identificar conceptos clave como la autonomía relacional, la privacidad mental, la subjetividad y la cognición distribuida. Estos conceptos han sido esenciales para entender cómo las tecnologías cognitivas no sólo extienden nuestras capacidades, sino que también condicionan nuestras decisiones, valores y relaciones. Este análisis teórico ha servido como base para articular los principios éticos propuestos en el trabajo, subrayando la necesidad de un enfoque que respete y proteja estas dimensiones humanas fundamentales. Sin embargo, la exploración de estos

fundamentos también ha revelado la complejidad de estas interacciones, destacando la importancia de realizar investigaciones futuras que profundicen en la relación entre la mente extendida y las dinámicas sociales y culturales.

En segundo lugar, el examen crítico del impacto de las tecnologías cognitivas ha puesto de manifiesto una serie de riesgos y desafíos éticos, como la manipulación algorítmica, la erosión de la privacidad mental y la pérdida de autonomía. Estos riesgos han sido contextualizados dentro de la teoría de la TME, lo que ha permitido evidenciar cómo las tecnologías pueden influir en los procesos cognitivos internos y las relaciones interpersonales. Aunque este análisis ha proporcionado una base para identificar los problemas éticos más relevantes, su alcance ha estado limitado por la falta de estudios empíricos que analicen casos concretos de tecnologías cognitivas en funcionamiento. Incorporar ejemplos prácticos y estudios de caso en investigaciones futuras podría fortalecer aún más las conclusiones derivadas de este trabajo.

En tercer lugar, la evaluación de los marcos éticos —consecuencialismo, deontología y ética de la virtud— ha demostrado que, si bien aportan perspectivas valiosas, son insuficientes para abordar la complejidad relacional y distribuida de las tecnologías cognitivas. Este análisis ha subrayado las limitaciones del consecuencialismo para considerar valores cualitativos como la privacidad y la autonomía, la rigidez de la deontología frente a dilemas tecnológicos contextuales, y la falta de herramientas sistemáticas en la ética de la virtud para guiar decisiones en entornos opacos y dinámicos. A partir de esta evaluación, se ha justificado la necesidad de un marco ético alternativo que integre estas perspectivas, pero que también responda a las especificidades de la relación humano-tecnológica. No obstante, un desafío pendiente es desarrollar herramientas prácticas que permitan operacionalizar este marco en el diseño y la regulación tecnológica.

Finalmente, el marco ético relacional de la responsabilidad propuesto en este trabajo representa un esfuerzo por responder a los desafíos identificados:

1. El principio de claridad y autonomía busca garantizar que los usuarios mantengan su autoridad epistémica y su capacidad crítica frente a las tecnologías cognitivas. Este principio responde al desafío de proteger la privacidad mental, la erosión de la autonomía y la dependencia opaca.

2. El principio de responsabilidad interdependiente y justicia refleja las dinámicas relacionales y enfatiza la distribución de la cognición en redes híbridas. Este principio

aborda la necesidad de equilibrar responsabilidades entre diferentes actores, evitando dinámicas de dominación tecnológica que comprometan la autonomía.

3. El principio de prospectiva ética y beneficencia introduce una visión prospectiva para anticipar y mitigar riesgos, mientras fomenta valores positivos como la equidad y la sostenibilidad. Este enfoque se complementa con estrategias como el codiseño, que asegura la inclusión de múltiples perspectivas en el desarrollo tecnológico.

Los principios propuestos reflejan una visión integradora que combina las demandas éticas actuales con la necesidad de anticipar riesgos futuros. Sin embargo, el marco también enfrenta limitaciones: su implementación depende de regulaciones, prácticas de diseño y mecanismos de supervisión ética que aún están en desarrollo. Además, la aplicación efectiva de estos principios requerirá una mayor colaboración interdisciplinaria entre diseñadores, usuarios, filósofos, legisladores y otras partes interesadas.

Una de las líneas de investigación más relevantes que surge de este trabajo se centra en el desarrollo de un marco prospectivo para la anticipación y mitigación de riesgos futuros, basado en los principios de este marco ético. Este enfoque busca integrar herramientas prospectivas que permitan anticipar impactos éticos y sociales antes de la implementación tecnológica. Estas herramientas incluirían talleres de codiseño, simulaciones éticas y modelos de visualización colaborativa, diseñados para explorar escenarios futuros y evaluar proactivamente sus posibles consecuencias. Esto fomentaría la participación activa de usuarios, diseñadores y expertos en ética, garantizando la inclusión de múltiples perspectivas en el proceso de diseño.

Además, estas herramientas prospectivas funcionarían como referentes conceptuales, ayudando a los diseñadores a alinear sus decisiones con los derechos fundamentales. Se complementarían con indicadores interdisciplinarios que integren conocimientos de la neurociencia, la psicología y la sociología, ofreciendo una base para evaluar las implicaciones futuras. Ejemplos de estas herramientas prácticas podrían incluir matrices de evaluación ética, indicadores adaptables a contextos diversos y modelos de simulación capaces de analizar el impacto a largo plazo de las tecnologías en las dinámicas cognitivas y relacionales.

Por último, se sugiere explorar mecanismos de supervisión ética integrados en las tecnologías, permitiendo un monitoreo y retroalimentación continuo, auditorías periódicas y análisis adaptativos basados en datos contextuales. Estos mecanismos podrían adaptarse de forma dinámica a los cambios culturales, sociales o tecnológicos, asegurando que las

tecnologías cognitivas se mantengan alineadas con los principios fundamentales y sean capaces de ajustarse a los contextos específicos de aplicación.

En síntesis, este trabajo no solo ha planteado una base teórica sólida para comprender los retos éticos derivados de las tecnologías cognitivas desde la perspectiva de la TME, sino que también ha propuesto un marco ético relacional de la responsabilidad como una respuesta integradora y proactiva. Este marco se posiciona como un punto de partida para el diseño de tecnologías que respete los principios marcados en este trabajo. Sin embargo, el camino hacia su implementación efectiva es un desafío abierto que requiere un compromiso interdisciplinario, colaborativo y adaptativo. Las herramientas prospectivas, los estudios de caso empíricos y la integración de dinámicas culturales y sociales serán esenciales para garantizar que las tecnologías futuras no solo respondan a los desafíos éticos actuales, sino que también anticipen y minimicen riesgos emergentes. Así, se reafirma la urgencia de seguir investigando y desarrollando soluciones que equilibren innovación, equidad y sostenibilidad en un mundo cada vez más interconectado.

5. Bibliografía y Webgrafía

5.1. Bibliografía

Apud, I. (2014). ¿La mente se extiende a través de los artefactos? Algunas cuestiones sobre el concepto de cognición distribuida aplicado a la interacción mente-tecnología. *Revista de Filosofía (Madrid)*, 39(1), 137-161.

Albrechtslund, A. (2007). Ethics and technology design. *Ethics and Information Technology*, 9(1), 63-72. https://doi.org/10.1007/s10676-006-9129-8

Becker, R., Chokoshvili, D., Comandé, G., Dove, E. S., Hall, A., Mitchell, C., Molnár-Gábor, F., Nicolás, P., Tervo, S., & Thorogood, A. (2023). Secondary use of personal health data: When is it "further processing" under the GDPR, and what are the implications for data controllers? *European Journal of Health Law, 30*(1), 129–157. https://doi.org/10.1163/15718093-bja10094

Blitz, M. J. (2010). Freedom of thought for the extended mind: Cognitive enhancement and the constitution. *Wisconsin Law Review*, 2010(4), 1049–1118.

Bratman, M. E. (2000). Reflection, planning, and temporally extended agency. *The Philosophical Review*, 109(1), 35–61.

Braudo-Bahat, Y. (2017). Towards a relational conceptualization of the right to personal autonomy. *American University Journal of Gender, Social Policy & the Law, 25*(2), Disponible en http://digitalcommons.wcl.american.edu/jgspl/vol25/iss2/1

Broncano, F. (2006). Sujeto y subjetividad en la mente extensa. *Revista de Filosofía,* 31(2), 109–133.

Burrell, J. (2015). How the machine 'thinks:' Understanding opacity in machine learning algorithms. *Big Data & Society, 3*(1). https://doi.org/10.2139/ssrn.2660674

Carter, J. A., & Palermos, S. O. (2016). Is having your computer compromised a personal assault? The ethics of extended cognition. *Journal of the American Philosophical Association.* https://doi.org/10.1017/apa.2016.28

Clark, A., & Chalmers, D. (1998). The extended mind. *Analysis*, 58(1), 7-19.

Clark, A., & Chalmers, D. (2011). La mente extendida (E. Aladro Vico, trad.). *CIC: Cuadernos de información y comunicación*, 16, 15-28.

Clowes, R. W., Smart, P. R., & Heersmink, R. (2024). The ethics of the extended mind: Mental privacy, manipulation and agency. En J.-H. Heinrichs, B. Beck, & O. Friedrich (Eds.), *NeuroProsthEthics: Ethical implications of applied situated cognition* (pp. 13–35). J. B. Metzler.

Cortina, A. (2024). ¿Ética o ideología de la inteligencia artificial? El eclipse de la razón comunicativa en una sociedad tecnologizada. Paidós.

Gilligan, C. (2013). *La ética del cuidado*. Fundació Víctor Grífols i Lucas.

Gunkel, D. J. (2020). Mind the gap: Responsible robotics and the problem of responsibility. *Ethics and Information Technology, 22*, 307–320. https://doi.org/10.1007/s10676-017-9428-2

Delacroix, S., Veale, M. (2020). Smart technologies and our sense of self: Going beyond epistemic counter-profiling. En M. Hildebrandt & K. O'Hara (Eds.), *Life and the Law in the Era of Data-Driven Agency* (pp. 80–99). Edward Elgar Publishing.

Dennett, D. (1996). *Kinds of minds: Towards an understanding of consciousness*. Basic Books.

Donald, M. (1991). *Origins of the modern mind*. Harvard University Press.

Domènech, M., Tirado, F. J. (s.f.). *La teoría del actor-red: Una aproximación simétrica a las relaciones entre ciencia, tecnología y sociedad*. FUOC.

Dworkin, G. (1981). The concept of autonomy. *Grazer Philosophische Studien*, 12, 203–213.

Floridi. L. (2024). *Ética de la inteligencia artificial. Principios retos y oportunidades.* Herder.

Farina, M., Lavazza, A. (2022). Incorporation, transparency, and cognitive extension: Why the distinction between embedded and extended might be more important to ethics than to metaphysics. *Philosophy & Technology, 35*(10). https://doi.org/10.1007/s13347-022-00508-4

Frankfurt, H. G. (1971). Freedom of the will and the concept of a person. *Journal of Philosophy*, 68(1), 5–20.

Friedman, M. (1997). Autonomy and social relationships: Rethinking the feminist critique. En D. T. Meyers (Ed.), *Feminists rethink the self* (pp. 40–61). Boulder, Colo.: Westview Press.

Friedman, B., Kahn, P. H., & Borning, A. (2002). Value sensitive design: Theory and methods. *University of Washington*.

Hanson, F. A. (2009). Beyond the skin bag: On the moral responsibility of extended agencies. *Ethics and Information Technology, 11*(1), 91–99.

Heinrichs, J.-H. (2018). *Neuroethics, cognitive technologies and the extended mind perspective.* Neuroethics, 14(1), 59–72. https://doi.org/10.1007/s12152-018-9365-8

Hutchins, E. (2000). *Cognition in the wild.* MIT Press.

Ihde, D. (1977). *Experimental phenomenology: An introduction*. Putnam.

Jonas, H. (1995). *El principio de responsabilidad: Ensayo de una ética para la civilización tecnológica.* Herder Editorial.

Kant, I. (2019). *Fundamentación para una metafísica de las costumbres.* Alianza Editorial.

Kirsh, D. (2006). Explaining artifact evolution. En L. Malafouris (Ed.), *Cognitive life of things: Recasting the boundaries of the mind*. McDonald Institute for Archaeological Research.

Latour, B. (2012). Primera carta. En *Cogitamus: Seis cartas sobre las humanidades científicas* (pp. 14–39). Paidós.

Levy, N. (2007). *Rethinking neuroethics in the light of the extended mind thesis*. *The American Journal of Bioethics, 7*(9), 3-11. https://doi.org/10.1080/15265160701518466

Llinares, J. (2018). Hacia una ética para el mundo tecnológico. *ArtefaCToS. Revista de estudios de la ciencia y la tecnología*, *7*(1), 99-120

Lynch, M. P. (2013). *Brief of Michael P. Lynch as amicus curiae in support of the plaintiffs*. Caso No. 13-cv-03994 (WHP), United States District Court, Southern District of New York. Disponible en https://www.aclu.org/legal-document/aclu-v-clapper-amicus-brief-michael-p-lynch-philosophy-professor-university

McCarthy-Jones, S. (2019). The autonomous mind: The right to freedom of thought in the twenty-first century. *Frontiers in Artificial Intelligence*, 2, 19. https://doi.org/10.3389/frai.2019.00019.

Menary, R. (2010). Cognitive integration and the extended mind. En R. Menary (Ed.), *The Extended Mind* (pp. 267-288). MIT Press. https://doi.org/10.7551/mitpress/9780262014038.003.0010

Mill, J. S. (1861/1998). *Utilitarianism.* Oxford University Press.

Nedelsky, J. (1989). Reconceiving autonomy: Sources, thoughts, and possibilities. *Yale Journal of Law & Feminism*, 1(1), 5.

Nedelsky, J. (2011). *Law's Relations: A Relational Theory of Self, Autonomy, and Law.* DOI:10.1093/acprof:oso/9780195147964.001.0001

Palermos, S. O. (2023). Data, metadata, mental data? Privacy and the extended mind. *AJOB Neuroscience*, 14(2), 84-96. https://doi.org/10.1080/21507740.2022.2148772.

Pérez Chico, D. (2010). Los límites de la tesis de la mente extendida: Agencia, autonomía y autoridad epistémica. *Factótum*, 7, 62–75.

Reiner, P. B., & Nagel, S. K. (2017). Technologies of the extended mind: Defining the issues. In J. Illes (Ed.), *Neuroethics: Anticipating the Future* (pp. 108–122). Oxford University Press.

Ronell, A. (1989). *The telephone book: Technology, schizophrenia, electric speech*. University of Nebraska Press.

Shen, F. X. (2013). Mind, body, and the criminal law. *Minnesota Law Review*, 97, 2036–2175.

Vega, J. (2005). Mentes híbridas: Cognición, representaciones externas y artefactos epistémicos. *AIBR. Revista de Antropología Iberoamericana*.

Verbeek, P. P. (2011). *Moralizing technology: Understanding and designing the morality of things.* University of Chicago Press.

Vold, K., Hernández-Orallo, J. (2019). AI extenders: The ethical and societal implications of humans cognitively extended by AI. *Proceedings of the AAAI Conference on Artificial Intelligence.* https://doi.org/10.48550/arXiv.2003.04881

Vold, K., Hernández-Orallo, J. (2022). *AI extenders: The ethical and societal implications of humans cognitively extended by AI*. Leverhulme Centre for the Future of Intelligence, University of Cambridge, & Universitat Politècnica de València.

Yáñez González, J. E. (2021). La responsabilidad humana: Un análisis de la ética de la responsabilidad de Hans Jonas. *Revista de Filosofía, Facultad de Estudios Teológicos y Filosofía UCSC, 20*(2), 89–109.

5.2. Webgrafía

Parlamento Europeo y Consejo de la Unión Europea. (2024). Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y se modifican varios reglamentos y directivas. Diario Oficial de la Unión Europea, 1689. Disponible en http://data.europa.eu/eli/reg/2024/1689/oj.

Sinnott-Armstrong, W. (2021). Consequentialism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2021 ed.). Stanford University. https://plato.stanford.edu/entries/consequentialism/