



INTELIGENCIA ARTIFICIAL Y ÉTICA DISCURSIVA.

El problema de la imitación de la argumentación

Artificial intelligence and discourse ethics. The problem of mock argumentation

Künstliche Intelligenz und Diskursethik. Das Problem der Nachahmung von
Argumentation

Andrés Crelier

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina;
Universidad Nacional de Mar del Plata, Mar del Plata, Argentina
ORCID: <https://orcid.org/0000-0002-3845-7669>

andrescrelier@gmail.com

Recibido: 21-12-2023

Aceptado: 02-04-2024

Andrés Crelier es Profesor en la Universidad Nacional de Mar del Plata e Investigador en Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET, Argentina). Ha escrito libros y artículos en diversas áreas de la filosofía contemporánea, como la hermenéutica, la filosofía trascendental y la ética discursiva. Actualmente, su interés se centra en la relación entre pensamiento y lenguaje, enfocados desde la filosofía analítica y, en especial, desde la relación con la etología cognitiva.

Resumen

El artículo trata el problema que representa para la ética discursiva la imitación de la argumentación por parte de sistemas artificiales. La primera parte explica el problema, señalando que la ética discursiva ha considerado que la manifestación usual de la capacidad de argumentar representa una condición suficiente para otorgar estatus

racional y moral a la criatura que argumenta. Esto obligaría a incluir dentro de la esfera de la racionalidad y la moralidad a los modelos de lenguaje que imitan los productos de una argumentación. La segunda parte desarrolla un modo de tratar este problema a partir de una analogía con el problema de cómo constatar la capacidad de sintiencia en criaturas no humanas y en sistemas artificiales. La tercera parte propone una solución teórica al problema, consistente en tomar en cuenta no sólo las manifestaciones discursivas de la argumentación sino las capacidades mentales, sociales y agenciales que esta capacidad involucra.

Palabras clave: *Ética Discursiva; Argumentación; Inteligencia artificial; Sintiencia.*

Abstract

The article addresses the problem that the imitation of argumentation by artificial systems represents for Discourse Ethics. Firstly, it points out that the latter has considered that the usual manifestation of the ability to argue represents a sufficient condition to grant rational and moral status. According to this, one should include those language models that imitate the products of an argument within the sphere of rationality and morality. Secondly, it develops a way of dealing with this problem based on an analogy with the problem of how to verify the capacity for sentience in non-human creatures and in artificial systems. Thirdly, it proposes a theoretical solution to the problem, which takes into account not only the discursive manifestations of argumentation but also the mental, social and agential capacities that this capacity involves.

Keywords: *Discourse Ethics; Argumentation; Artificial intelligence; Sentience.*

Zusammenfassung

Der Artikel befasst sich mit dem Problem, das die Nachahmung von Argumentation durch künstliche Systeme für die Diskursethik darstellt. Der erste Teil erklärt das Problem und weist darauf hin, dass die Diskursethik davon ausgegangen ist, dass die übliche Manifestation der Argumentationsfähigkeit eine ausreichende Bedingung darstellt, um dem Argumentierenden einen rationalen und moralischen Status zu verleihen. Dies würde dazu führen, dass Sprachmodelle, die die Produkte eines Arguments imitieren, in den Bereich der Rationalität und Moral einbezogen werden. Der zweite Teil entwickelt einen Weg, mit diesem Problem umzugehen, basierend auf einer Analogie zum Problem, wie die Empfindungsfähigkeit bei nichtmenschlichen Lebewesen und in künstlichen Systemen überprüft werden kann. Der dritte Teil schlägt eine theoretische Lösung des Problems vor, die darin besteht, nicht nur die diskursiven Erscheinungsformen des Arguments zu berücksichtigen, sondern auch die mentalen, sozialen und handelnden Fähigkeiten, die diese Fähigkeit mit sich bringt.

Schlüsselwörter: *Diskursethik; Argumentation; Künstliche Intelligenz; Empfindungsvermögen.*

Planteo del problema: la argumentación imitada

La ética discursiva clásica, tal como ha sido propuesta en su etapa inicial por Jürgen Habermas y especialmente por Karl-Otto Apel, considera que la capacidad de argumentar es un indicio seguro de racionalidad y, según su orientación kantiana, de lo que significa ser una persona en sentido moral (Apel, 1995; 1998). A mi modo de ver, esta tesis general no excluye en principio la posibilidad de que haya criaturas con estatus moral que no tengan esa capacidad, tal como sugieren muchas de nuestras intuiciones sobre niños pequeños y una variedad de animales no humanos. Por esa razón, se podría sostener que la capacidad de argumentar representa para la ética discursiva una condición suficiente, aunque no necesaria, de la moralidad. En todo caso, ha sido la presencia de esa capacidad lo que la ética discursiva ha tenido en cuenta como base para el desarrollo de su propuesta teórica.

Ahora bien, es posible hacer una distinción entre la posesión de una capacidad y su ejercicio, en el sentido de que una misma capacidad se puede manifestar de diferentes maneras. Así, quien posee la capacidad de argumentar puede manifestarla de diferente modo y mediante diferentes vehículos signícos, tales como el sonido en la expresión oral, el signo en la escritura o los gestos en un lenguaje de señas. A su vez, puede ejercer esta habilidad en el marco de dos dimensiones temporales, la sincrónica en el habla oral o la diacrónica en la escritura. De este modo, podemos afirmar que la capacidad de argumentar no se observa tal como es “en sí misma” sino a través de sus manifestaciones o “indicadores”, presentes cuando la habilidad es efectivamente ejercida.

Como señalé al inicio, la ética discursiva considera que la presencia de esta capacidad, que se manifiesta a través de indicadores como los mencionados, es una prueba de racionalidad y de estatus moral. En esto, sigue una tradición que se remonta a Aristóteles. En efecto, la idea de que existen indicadores que muestran, a quien se los exhibe, que los individuos que los exhiben se encuentran insertos en el dominio de la racionalidad se encuentra en su *Metafísica*.

En esa obra, Aristóteles explora el mejor modo de mostrar la validez del principio de no contradicción, es decir, de un principio constitutivo de la propia racionalidad. Señala que el mero uso del lenguaje es un indicador, lo cual se pone de relieve en que su ausencia equivale a encontrarse frente a una criatura no racional, como un vegetal (Aristóteles, *Metafísica* 1006a, 10-15). En el seno de esta misma tradición, y con referencias explícitas a Aristóteles, la ética discursiva (al menos la apeliana) sostiene que la capacidad de argumentar es un indicador de racionalidad y también de moralidad. En desarrollos teóricos destinados a darle sustento de la propuesta ética, distintos representantes de esta posición han mostrado que, quien argumenta, realiza necesariamente presupuestos de racionalidad y moralidad (Apel, 1995; Crelier, 2010).

En este trabajo no abordaré la problemática de esta reconstrucción, que considero posible y válida, sino que me concentraré en otro problema teórico. Se trata de cómo determinar cuáles son los marcadores fiables de la argumentación. Este problema no ha sido considerado en cuanto tal en el debate filosófico abierto por la ética discursiva, ante todo porque se ha considerado que los indicadores usuales resultaban suficientes. Vale decir, no se consideraba que existía una brecha problemática entre la manifestación de la capacidad y la capacidad misma. El hecho de que una persona formulara un argumento en un debate “en vivo” ha sido implícitamente considerado como una muestra de su genuina capacidad de argumentar, y lo mismo ha sido presupuesto *mutatis mutandis* en el caso de la argumentación expresada por escrito, entre otros modos en que esta habilidad se manifiesta en las personas hablantes.

Ahora bien, la confiabilidad de los indicadores puede ponerse en duda a partir de desarrollos recientes de la inteligencia artificial en sistemas informáticos. Aunque podamos estar de acuerdo en que un adulto humano que argumenta oralmente, frente a una audiencia de otros adultos humanos, posee la capacidad genuina de hacerlo, y una planta que no lo hace no la posee, surge la pregunta acerca de si un sistema de inteligencia artificial que ofrece, a través por ejemplo de una pantalla, una secuencia de signos que tienen toda la apariencia de argumentos, exhibe con ello una capacidad argumentativa auténtica.

En efecto, la posibilidad de que los sistemas de inteligencia artificial exhiban resultados discursivos muy similares a los que ofrece la capacidad de argumentar está actualmente fuera de discusión. Los sistemas que imitan el lenguaje humano han llegado a ser muy flexibles y convincentes. Son capaces de responder de manera coherente — por escrito en una pantalla y también oralmente— a preguntas sobre conceptos que, como ha sido distintivo para nuestra tradición filosófica, se inician con la expresión “por qué”. Tomo como ejemplo un fragmento de conversación con el sistema ChatGPT que responde de manera coherente a una pregunta moral clásica: ¿Por qué robar está mal?:¹

“El acto de robar se considera moralmente incorrecto por varias razones fundamentales:

Violación de la propiedad: El robo implica tomar o usar algo que legítimamente pertenece a otra persona sin su consentimiento. Esto viola el derecho a la propiedad de esa persona y socava la confianza y la seguridad en la sociedad.

Conflicto con la ley: En la mayoría de las sociedades, el robo es ilegal y está castigado por la ley. Cometer un robo puede llevar a consecuencias legales graves, como la prisión o multas, lo que refuerza la idea de que es un comportamiento inaceptable.

Daño emocional y psicológico: Ser víctima de un robo puede causar un profundo estrés emocional y psicológico a la persona afectada. Puede sentirse violada, vulnerable y traumatizada, lo que demuestra el impacto negativo que tiene el robo en las personas.”²

¿Manifiesta esta coherente e informada respuesta una capacidad de auténtica racionalidad discursiva? ¿Estamos ante un discurso que contiene marcadores confiables de la capacidad de argumentar? Una respuesta negativa a preguntas de esta clase entra en conflicto con un supuesto de la ética discursiva, según el cual la razón suficiente más relevante para la inclusión moral es manifestación de una capacidad para argumentar, no las propiedades físicas o incluso psicológicas de la criatura que posee esa capacidad. Justamente, el poder igualador de la argumentación y su relevancia moral consisten en que debemos respetar los derechos de todos los argumentantes. En el ámbito humano, esto implica respetar a los que argumentan cualquiera sea su color de piel, grupo étnico,

¹ Se puede alegar aquí los sistemas de lenguaje actuales no argumentan sino que informan sobre argumentos, como sucede de hecho en el ejemplo seleccionado. En todo caso, esto es algo que usualmente hacemos las personas que argumentamos, y que podría ser fácilmente modificado por el diseño de estos sistemas para que la imitación se acerque más a un diálogo con una segunda persona sobre sus propias opiniones.

² El fragmento fue generado por el modelo de lenguaje en la plataforma <https://chat.openai.com/>.

edad o nacionalidad. Y no hay en principio razones para restringir este supuesto a quienes pertenecen a determinada especie. Todo esto favorece una respuesta positiva a las preguntas que inician este párrafo.

Pero si damos una tal respuesta positiva, en el sentido de que el discurso elaborado por el modelo de lenguaje es un indicador confiable de la capacidad de argumentar, entonces deberíamos —desde la ética discursiva— incluir a este sistema de inteligencia artificial dentro del ámbito de las “criaturas” con estatus moral. El problema es que esto iría contra intuiciones del (actual) sentido común. Pues suele pensarse que los sistemas informáticos que elaboran estos modelos del lenguaje no son verdaderas criaturas pensantes. En correspondencia con esto, tenemos la intuición de que resulta descabellado otorgar estatus racional y moral a un sistema informático.

Peor aún, resultaría anti-intuitivo reconocer la obligación de otorgarle estatus moral a un sistema de inteligencia artificial y no reconocer una clara obligación de otorgárselo —aunque sea de manera gradualista— a criaturas que no son capaces de argumentar o que han perdido esa capacidad. Aquí se incluyen infantes humanos pre-lingüísticos, personas que han perdido capacidades cognitivas básicas y muchas especies no humanas que, en general, consideramos dignas de trato moral. Ninguna de las criaturas mencionadas es capaz de responder a una pregunta “por qué” del modo en que lo ha hecho el sistema informático diseñado para conversar sobre una multiplicidad de temas de manera coherente, o en todo caso para simular una tal conversación.

A mi modo de ver, el teórico de la ética discursiva puede dar crédito a estas fuertes intuiciones morales y tomarlas como una motivación para discutir, con mirada crítica, si es posible incluir sistemas artificiales dentro de la esfera de la moralidad. En esta indagación crítica puede tomar en cuenta la señalada brecha entre la capacidad y su manifestación, en este caso, entre la capacidad de argumentar y el discurso coherente que presuntamente se manifiesta como su resultado. En este marco, puede sostener que, en el caso de la inteligencia artificial, los frutos de la aparente capacidad de argumentar no son en realidad indicadores confiables o suficientes de que estamos ante una capacidad genuina. Aquí puede hacer valer las mencionadas intuiciones sobre los sistemas artificiales, en el sentido de que la sospecha sobre los indicadores es a su vez

una sospecha sobre la fuente que los produce. Una vez establecidas estas sospechas, surge naturalmente la necesidad de proponer otros indicadores que confirmen, de manera independiente, que estamos ante una capacidad argumentativa auténtica y por ende ante una criatura con estatus moral.

Ahora bien, una objeción inicial contra esta solución es que parece prescindir de la manifestación de la argumentación, pues consiste en sostener que debemos contar con otras pruebas de la posesión de estatus racional, de modo que la manifestación de la argumentación no sería ya una razón suficiente de inclusión en el orden de la racionalidad y la moralidad. Pero esto parece ir contra el punto de partida ético discursivo, para la cual no existiría mejor indicación de que estamos ante una criatura racional que su capacidad para argumentar tal como usualmente esta se manifiesta, por ejemplo en una respuesta a la pregunta acerca de por qué está mal robar.

En este punto, la ética discursiva puede adoptar una posición intermedia —que yo mismo adoptaré—, insistiendo en que hay que tomar en cuenta los aspectos manifiestos de la argumentación, pero aceptando también que debemos buscar criterios adicionales para distinguir entre una argumentación genuina y otra que no lo es, en tanto sólo constituye una imitación de la argumentación. Ahora bien, esto requiere explicitar cuáles son esos indicadores adicionales de argumentación genuina. Para explicar esta opción más detenidamente, voy a desarrollar una analogía con un problema que guarda similitud con el que estoy formulando, el que se refiere a cómo constatar la presencia de la capacidad de sintiencia en criaturas no humanas y sistemas de inteligencia artificial.

Desarrollo del problema de los indicadores: analogía con la sintiencia

Con el objetivo de proponer una noción de argumentación ampliada, que podría ser una manera de tratar el problema planteado dentro del marco de la ética discursiva, voy a proponer la siguiente analogía. Así como distinguíamos más arriba entre la manifestación de la argumentación, por un lado, y la capacidad de argumentar, por el otro, se puede distinguir entre los indicadores de conciencia fenoménica, por un lado, y

la misma capacidad de conciencia fenoménica, por el otro. En ambos casos, existen marcadores más o menos confiables de una capacidad que está “por detrás” (lo cual debe tomarse en un sentido metafórico, pues en todo caso no existe una separación entre la capacidad ejercida y su ejercicio).

En el caso de la conciencia fenoménica, el debate gira en torno a cómo verificar su presencia genuina en diferentes criaturas no humanas o incluso en sistemas artificiales. Más específicamente, se trata de indagar capacidades de “sintiencia”, como la capacidad de sentir dolor y placer (entre otros afectos), y más concretamente aún, la discusión ha tenido en cuenta la capacidad de sentir dolor, ante todo por su clara relevancia ética. En ambos terrenos —sintiencia y argumentación— aquello que se manifiesta es sólo un marcador falible, y por ello precisamos determinar cuáles son los marcadores confiables de la capacidad en cuestión.

En este apartado seguiré de cerca la reciente presentación del problema de la sintiencia no humana en Andrews y Birch (2023) (ver también Browning y Birch, 2022). El punto de partida es la dificultad de contar con un reporte verbal que nos confirme (o no) la presencia de capacidades afectivas, ante todo en la investigación de especies no humanas. Sí tenemos, en cambio, la posibilidad de investigar aspectos fisiológicos y compararlos con la fisiología de los humanos, de quienes sabemos que sienten dolor y placer. Pero dada la diversidad de sistemas nerviosos, o incluso la existencia de animales con conductas complejas sin un sistema nervioso central como los pulpos, los indicadores de orden fisiológico resultan insuficientes por sí solos.

Esto lleva a atribuir una significación central a los indicadores conductuales de la capacidad de sintiencia más estudiada, el dolor. Entre estos se cuentan la atención de cuidado que la criatura manifiesta hacia un tejido dañado, la evitación de lugares nocivos y la preferencia de sitios donde hay drogas analgésicas. En base a estos indicadores, sumados en ocasiones a los anteriores o como complemento de los mismos, se ha empezado a considerar la posibilidad de que organismos invertebrados sientan dolor.

Ahora bien, ¿cómo sabemos que estos indicadores son fiables, dada la posibilidad de que algunos de ellos sean imitados por sistemas artificiales? En términos generales, los diferentes tipos de imitación de sistemas biológicos han mejorado con el tiempo,

abarcando diferentes aspectos. Así, se ha “emulado” el cerebro de algunos organismos relativamente simples en un ordenador y se está explorando en la actualidad el modo de hacerlo con organismos menos sencillos. También es posible imitar la conducta observable, ante todo la gestual, de una criatura humanoide que siente dolor.

Sin embargo, quizás la emulación artificial más lograda hasta el momento sea conductual y de naturaleza específicamente lingüística. Los modelos de lenguaje de la inteligencia artificial son capaces de ofrecer, como vimos más arriba, una manifestación signífica de una conversación auténtica. Y así como ofrecen respuestas verbales que imitan una argumentación, pueden muy bien ofrecer manifestaciones “externas”, de naturaleza también verbal, de sintiencia como dolor o placer. Tomando como base una gran cantidad de información, estos modelos extraen patrones y los usan para ofrecer los resultados de una conducta verbal sofisticada y sorprendente. Y es en principio factible una combinación de la imitación lingüística con la imitación fisiológica y gestual.

Ahora bien, Andrews y Birch señalan que estas emulaciones se desarrollan en sistemas muy diferentes a los organismos vivos, y que se trata de sistemas que han sido diseñados con el fin expreso de imitar la conducta de estos últimos. Así, los modelos de lenguaje han sido programados con la intención de inferir patrones a partir de una enorme cantidad de información e imitar con ello algunos de los productos de la conducta verbal. El hecho de que hayan sido diseñados para tal fin les quita pues valor como evidencia de que son marcadores fiables de sintiencia. De hecho, a raíz de este hecho los autores los descartan en cuanto tales. Andrews y Birch piensan que sería entonces conveniente contar con una teoría que tome en consideración una serie de casos que, según sabemos de manera fehaciente, son ejemplos genuinos de sintiencia. Estos casos serían los que ofrecerían marcadores fiables, que podríamos entonces volver a emplear en la investigación.

Pasemos ahora al caso de la argumentación. Puede considerarse aquí que los candidatos a indicadores fiables de la presencia de esta capacidad se encuentran en las distintas dimensiones del lenguaje que típicamente se usan en un intercambio argumentativo. Un repaso somero por estos aspectos deja en evidencia que resultan, en principio, imitables por un modelo artificial de lenguaje. A nivel pragmático, se

encuentra el uso de lenguaje natural de manera flexible y contextual. La capacidad de argumentar incluye la capacidad de identificar la relevancia de la información y responder por ejemplo a preguntas “por qué” en el seno de una conversación sobre un tema específico. Esto incluye la capacidad de ofrecer manifestaciones visuales y auditivas adecuadas, incluyendo tonos que marcan emociones.

A nivel sintáctico, se halla la capacidad para combinar los signos para producir unidades nuevas y más complejas, según reglas precisas que admiten ciertas combinaciones y no otras. A nivel semántico, encontramos la misma combinabilidad y productividad para producir unidades mayores con un significado coherente. A nivel semántico-inferencial, la capacidad referencial se liga con capacidad de establecer relaciones inferenciales pertinentes en el marco de argumentos complejos.

Sin volver a una ilustración de detalle, queda claro que los modelos de lenguaje son capaces de imitar de manera convincente los tres aspectos del lenguaje mencionados. Esto nos permite seguir adelante y discutir la analogía con el caso de la sintiencia, destacando también las diferencias entre ambos. Ante todo, existe una aparente ventaja para evaluar los marcadores de la argumentación, que no existiría con los marcadores de sintiencia. En este último caso, se trata de un acceso indirecto a partir de manifestaciones falibles. Así, mientras que los marcadores de dolor son necesariamente externos, los de la argumentación parecen ser una parte constitutiva de la capacidad investigada, incluso si aceptamos que existe una brecha entre la capacidad y su ejercicio. Más aún, estos “marcadores directos” incluyen una diversidad de aspectos (sintaxis, semántica, pragmática, inferencia).

Sin embargo, existe una gran desventaja en relación con el caso de la argumentación. Mientras que en el caso de la sintiencia algunos marcadores importantes son difícilmente imitables, en el caso de la argumentación los marcadores parecen ser todos imitables por los modelos de lenguaje. Así, mientras que la fisiología y la conducta flexible que manifiesta dolor son difíciles de imitar en su grano fino, los resultados de la capacidad de argumentar resultan —como hemos visto— claramente imitables por los recientes modelos de lenguaje.

Modos de tratar el problema

En relación a cómo constatar la presencia de sintiencia en un animal o sistema artificial, Andrews y Birch (2023) proponen lo siguiente. Creen que es preciso dejar de lado la idea de que los marcadores más clásicos alcanzan por sí solos y consideran necesario observar la arquitectura profunda de las criaturas que los exhiben. Esto los conduce a verificar la gran diferencia que existe entre los sistemas diseñados para imitar y aquellos que son fruto de la evolución natural. En principio, sólo en estos últimos podemos hallar ejemplos genuinos de conciencia fenoménica y de capacidades como la de sentir dolor y placer, pero no en los otros.

Con esto volvemos a la idea de tomar en cuenta aquellos marcadores de sintiencia en los organismos que son el fruto de una evolución natural. Dentro de este conjunto previamente delimitado de organismos, los marcadores clásicos vuelven a ser suficientes o adquieren al menos mayor relevancia, ya que la presencia de un marcador no se encuentra sujeto a la sospecha de que es producto de una imitación prevista por un diseñador (entendido este último como un diseñador intencional). Esto se refuerza con la idea de que, si contamos con una amplia cantidad de organismos que exhiben marcadores apropiados, podemos estudiar mejor la naturaleza de la sintiencia auténtica.

Esta estrategia tiene sus límites, pues se trata nuevamente de interpretar marcadores, que en cuanto tales no dejan de ser indicadores de sintiencia, sólo que ahora serían (al menos en parte) a prueba de imitación. Pero nada nos asegura de manera concluyente que algunos organismos que son el resultado de la evolución natural exhiban marcadores sin tener de hecho la capacidad de sintiencia. Si bien se pueden descartar *ex hypothesi* sistemas de inteligencia artificial, no es posible discriminar, entre los seres vivos, cuáles tienen conciencia fenoménica y cuáles no la tienen (Carruthers, 2019). Peor aún, nada impide que se puedan diseñar sistemas artificiales que no sólo exhiban marcadores sin contar con la capacidad, sino que los posean de manera genuina, al margen de que los tengan para engañar.

Pero sin avanzar con consideraciones que refuerzan una mirada escéptica sobre la sintiencia, y quedándonos con la útil diferenciación entre organismos que son fruto de la evolución natural y sistemas artificiales que han sido diseñados con fines específicos, ¿qué tipo de conclusiones se pueden extrapolar al caso de la argumentación? A mi modo de ver, resulta provechoso volver la mirada sobre la “estructura subyacente” de la criatura que siente dolor o, en este caso, del sistema que ofrece evidencia de que argumenta racionalmente.

Podemos sostener que la argumentación genuina, en cuanto capacidad, está constitutivamente ligada con otras capacidades. Por un lado, con facultades mentales sofisticadas como la memoria, la capacidad inferencial, la identidad personal, entre otras; y, por el otro, con capacidades de agencia que deben guardar coherencia práctica con la argumentación, y que incluyen la relación intersubjetiva con otros argumentantes en una “comunidad de comunicación” (según la expresión de Apel). En esto podemos fundar la tesis de que, en ausencia de estas facultades mentales y agenciales, no estamos ante la manifestación de una argumentación genuina, incluso cuando un modelo de lenguaje nos ofrezca refinados discursos y respuestas coherentes e informadas a nuestras preguntas.

Si se aceptan estas consideraciones, constatar la presencia de una diversidad de marcadores de la argumentación, cuando son tomados meramente como productos discursivos, no nos ofrece una condición suficiente para la atribución de la capacidad de argumentar (y por ende del estatus de racionalidad y moralidad que, según la ética discursiva, se encuentran ligados con esta capacidad). Esto supone que, ante la apariencia de argumentación, se requieren pruebas adicionales e independientes de que estamos ante una criatura con una racionalidad discursiva genuina.³

Un punto importante, si seguimos esta opción teórica, es que estas pruebas de la capacidad genuina no resultan independientes de la argumentación manifiesta tomada

³ Resulta pertinente aclarar que esta sugerencia de una investigación ampliada, que requiere ahora pruebas independientes que van más allá de la presencia de los resultados de la conducta verbal argumentativa, se encuadra en la búsqueda de condiciones suficientes de la racionalidad y de la moralidad. En particular, nada se afirma acerca del estatus moral de criaturas que carecen de la capacidad de argumentar, como los niños pequeños y algunos animales no humanos.

como producto. Las capacidades mentales y de agencia que he reseñado a manera de esbozo tienen que poder ejercerse y dar lugar a los productos adecuados, tal como podemos exigir de cualquier otra capacidad. Se trata de investigar una noción “extendida” de argumentación, que incluye capacidades subyacentes y adyacentes (mentales y agenciales) que hacen que una argumentación sea genuina en todos sus aspectos.

Esto supone que la ética discursiva “clásica” debería estar dispuesta a una suerte de ampliación, al menos en relación con una mirada restringida sobre la argumentación como un producto discursivo con ciertas características. En algún sentido, esta ampliación sigue la propuesta inicial de que la argumentación supone una comunidad de argumentación. Pero en otro sentido, se tienen en cuenta facultades y propiedades de los agentes que abarcan incluso al plano evolutivo.

Ya no es la capacidad manifiesta, en cuanto a los frutos o resultados de la argumentación, la base principal de la atribución de estatus moral, sino esta capacidad en relación con una serie de capacidades constitutivamente ligadas con ella. La racionalidad argumentativa, que indica estatus moral, sería pues la capacidad de un agente vivo, fruto de la evolución natural, que tiene capacidades mentales complejas, que vive en un medio social y que comparte un sistema de comunicación con otros argumentantes. Todo esto debe poder manifestarse no sólo en la conducta verbal sino también en la acción racional.

Referencias

- Apel, K.-O. (1995). *Teoría de la verdad y ética del discurso*. Barcelona: Paidós.
- Apel, K. O. (1998). *Auseinandersetzungen in Erprobung des transzendentalpragmatischen Ansatzes*. Frankfurt am Main: Suhrkamp.
- Aristóteles (2000). *Metafísica*. Madrid: Gredos.
- Andrews, K. y J. Birch (2023). What has feelings? *Aeon*. Disp. en <https://aeon.co/essays/to-understand-ai-sentience-first-understand-it-in-animals>
- Browning, H. y J. Birch (2022). Animal sentience. *Philosophy compass*, 17(5), e12822.

- Carruthers, P. (2019). *Human and animal minds: The consciousness questions laid to rest*. Oxford University Press.
- Crelier, A. (2010). *De los argumentos trascendentales a la hermenéutica trascendental*. La Plata: Edulp.