



**A PROPÓSITO DE UN TRABAJO DE GUNTER TEUBNER SOBRE LA
PERSONIFICACIÓN CIVIL DE LOS AGENTES DE INTELIGENCIA ARTIFICIAL
AVANZADA***

(Digitale Rechtssubjekte? Zum privatrechtlichen Status autonomer Softwareagentem, Archiv für die Civilistische Praxis, 218, 2018, pags. 155-205).

Angel Carrasco Perera**
*Catedrático de Derecho Civil
Centro de Estudios de Consumo
Universidad de Castilla-La Mancha*

Fecha de publicación: 11 de enero de 2019

1. Gunther Teubner es, *meiner Meinung nach*, el más esclarecido de los teóricos del Derecho vivos de Alemania. Aunque su producción es plural (incluso con incursiones en el Derecho civil “puro”), su obra más propia se construye sobre fundamentos de las teorías de sistemas y a partir de una interiorización jurídica de la concepción de Luhmann sobre el intercambio social como una acción comunicativa. El sustrato del desarrollo argumental que lleva a cabo el autor en el presente estudio puede ser adecuadamente compendiado en estas palabras de Lorentzen, que Teubner hace suyas: “*Los déficits de competencia psicosomática de los no-humanos quedan sobradamente compensados con la inteligencia distribuida de los sistemas sociales*”. El software inteligente al que Teubner dedica su consideración es el agente de Inteligencia Artificial capaz de producir *outputs* no preprogramados, capaces de aprender a partir del suministro de datos y susceptibles de elegir en situaciones de incertidumbre. Utilizando un término luhmanniano del que Teubner hace uso en otras obras (pero que aquí no emplea), se trataría de agentes algorítmicos “autopoiéticos” (cfr. también, en el mismo volumen, Marc Amstutz, *Dateneigentum. Funktion und Form*, pags 504-508).

* Trabajo realizado en el marco de la Ayuda del Programa Estatal de Fomento de la Investigación Científica y Técnica de Excelencia (Subprograma Estatal de Generación de Conocimiento) del Ministerio de Economía y Competitividad, otorgada al Grupo de investigación y Centro de investigación CESCO, Mantenimiento y consolidación de una estructura de investigación dedicada al Derecho de consumo, de la UCLM, ref. DER2014- 56016-P, del que soy investigador principal.

** <https://orcid.org/0000-0003-3622-2791>



2. La concepción del autor parte del postulado fundamental de que, o el Derecho civil se decide a reconocer a los *Software-Agentes* un cierto status subjetivo de *e-personas* que le cualifique como un “actores responsables”, y no mero *instrumentos* de humanos o de organizaciones de humanos, o se incrementará progresivamente el número de “accidentes” que no podrán imputarse a nadie a título de responsabilidad. Pero no propone una plena subjetividad de estos agentes. Esta capacidad jurídica “general” sólo tendría sentido en un estado hoy inexistente en el que la sociedad y el Derecho estuvieran dispuestos a atribuir y asignar a estos agentes recursos económicos propios con los que pudieran perseguir finalidades autónomas de lucro. De momento no es el caso.

3. Según el autor, un *Software-Agente* capaz de producir comunicación intersubjetiva no completamente preprogramada (la última generación en Inteligencia Artificial, *Deep learning*), capaz de estrategias de respuesta a las acciones externas y hábil para optimizar entre varios fines, ha de ver reconocida una suerte de capacidad jurídica, siquiera limitada, en la medida en que actúa con autonomía en el intercambio social comunicativo; no importa si este agente “no comprende” la subjetividad de su interlocutor humano – no existe la “doble contingencia” comunicativa del intercambio entre humanos- , bastando que el “mensaje de respuesta” se acomode al mensaje de pregunta. Propone huir de la imagen antropomorfista; su modelo no debe ser la persona física, sino las personas jurídicas, puras organizaciones y redes, también sin autoconsciencia ni voluntad subjetiva, que sin embargo actúan como centros de imputación de decisiones y conductas jurídicamente relevantes, como *cadena de decisiones*. Como las comunidades de propietarios, las sociedades no corporativizadas o las asociaciones no inscritas, se encuentran aquéllos en este estado de purgatorio social en el que pasan un tiempo hasta ser reconocidas como sujetos de Derecho. *Procesos comunicativos* que alcanzan un grado suficiente de *metabolización autónoma* del flujo de información social como para que el Derecho les reconozca una determinada capacidad de actuación jurídica y que son lo suficientemente complejos como para no satisfacerse con una descripción fiscalista de sus procesos de producción de *outputs* informacionales. Eliminemos del debate el grado inalcanzable de la autoconsciencia y sustituyamos “inteligencia” por “comunicación”. Llegamos a ese punto en que estos agentes puedan *tomar decisiones sobre un fondo de incertidumbre*, y ya es preciso en este instante atribuirles una responsabilidad propia de sujeto de Derecho. Esto sería así incluso si se diera por buena la propuesta radical de algún autor como Zech, según la cual es por principio un acto ilícito el empleo de algoritmos que permitan un desarrollo autónomo. Otros también han hablado en Alemania de la radical inconstitucionalidad de hacer convivir (“enfrentar”) a humanos con entidades autónomas de esta clase.



4. Frente a la concepción casi unánime en Alemania, que rehúsa dar importancia al fenómeno en sede de negocio jurídico – ya que las declaraciones negociales del *Software-Agente* se imputarían directamente a la persona física que lo explota económicamente / lo construye / lo programa, incluso por exigencias constitucionales relativas a la dignidad del ser humano-, Teubner recuerda el contenido de la sección 14 de la Uniform Electronic Transactions Act de los EEUU, que reza de esta forma: “*A contract may be formed by the interaction of electronic agents of the parties, even if no individual was aware of or reviewed the electronic agents’ actions or the resulting terms of the agreement*”. El autor propone reconocer a estos agentes al menos la capacidad jurídica necesaria para imputar la declaración negocial al principal *a título de representante*, mayormente para aislar al agente humano de la responsabilidad que conforme a Derecho civil no puede imputarse a un principal cuando el agente opera como un *falsus procurator* y al primero no se le puede atribuir la infracción de un deber de cuidado. En lo que respecta a la responsabilidad por incumplimiento contractual, si bien es ingeniosa la respuesta del autor, en términos prácticos no va más allá de lo que se llegaría negando toda imputación autónoma al agente robótico y predicarla directamente del gestor humano; porque en términos de responsabilidad contractual, es lo mismo que el agente humano responda del incumplimiento objetivamente por el riesgo de emplear determinados recursos para el cumplimiento o decir que el agente humano responde objetivamente de la conducta del agente artificial en cuanto ésta no es más que una conducta subordinada de *auxiliares de cumplimiento*, de los que se responde objetivamente (art. 1596 CC). No tiene importancia que en la dogmática jurídica se exija al menos culpa del auxiliar (en este caso, del autómeta), porque de hecho en materia contractual la culpa (*responsability*) se identifica con la simple ilicitud de la conducta de cumplimiento (*accountability*) y no aporta criterios adicionales para discriminar la responsabilidad. En el ámbito contractual da, pues, igual si el auxiliar en el cumplimiento es autónomo o si endosa la responsabilidad al principal como mero instrumento de cumplimiento.

5. En el ámbito extracontractual, el gestor del autómeta no estaría sujeto a una responsabilidad directa y primaria de naturaleza objetiva sobre la base del riesgo creado por el empleo de Inteligencia Artificial autónoma, sino a una responsabilidad vicaria por hecho de otro, incardinable en nuestro art. 1903 CC. Como en la responsabilidad contractual, es el propósito del autor encontrar un modo de no exponer directamente al principal a una suerte de responsabilidad absoluta por decisiones “ajenas” que él no ha podido controlar ni prever. Teubner entiende que no se puede hablar de una responsabilidad objetiva por riesgo de empleo de autómetas inteligentes, porque el agente humano creador del riesgo no debe responder de los daños (incluso lícitos) que realizan el riesgo típico, sino sólo de la conducta “defectuosa” del *Software-Agente*: “*el criterio de imputación no es el empleo de un objeto creador de un riesgo extraordinario, sino la*



conducta contraria a Derecho del algoritmo, que el principal ha empleado correctamente en su servicio". Para que el Software-agente pueda elevarse de esta norma a un nivel de sujeto capaz de imputación de responsabilidad, el aseguramiento obligatorio es una medida insoslayable: *"en este sentido, las soluciones asegurativas comportan un equivalente funcional de una subjetividad jurídica plena"*.

6. En la parte final de su trabajo el autor se entrega a interesantes reflexiones filosóficas – a las que el Derecho positivo hoy por hoy no puede secuenciar- relativas a la posibilidad de construir subjetividades y responsabilidades distribuidas en las redes multi-agentes entre humanos y no humanos, en la que sea posible identificar decisiones pero no actores a los que imputar su autoría, emergiendo de esta forma la necesidad de subjetivizar-responsabilizar las decisiones mismas y no los titulares de las misma. *"Aquí abandona el Derecho definitivamente la perspectiva de los actores, porque ya no se trata más de actores individuales o colectivos sino de decisiones arriesgadas. Las cadenas de acciones devienen en sí mismas responsables, sin preocuparse de su cualidad como unidades de conjunción de voluntades"*. El autor se ayuda con palabras elocuentes de una cita escogida de Hildebrandt: *"Because the agents may be distributed on and possibly mobile [sic] between different hardware applications and because as a multiagent system is capable of changing shape (plymorphous), it is not always easy to identify where the emerging agent is located and what is and is not a part of it at any point of time. However, insofar as the emergent behaviours of the system allow its identification as a unity of action, it can be qualified as an agent, whatever the underlying embodiment"*.

7. Para que estas propuestas tengan sentido, y para que existan institucionalmente islas de decisores autónomos que absorban la responsabilidad de sus decisiones y acciones sin trasladarlas *uno ictu* a los humanos, es preciso crear correspondientemente patrimonios de responsabilidad adscritos a tales decisores autónomos, que en el estado presente de la ciencia empezarán por ser patrimonios fundacionales personificados y adscritos a un fin, patrimonios sobre los que los autómatas no titulen derechos de propiedad (un *peculium* "adscrito" a un *alieni iuris*, como nos enseñaron los clásicos). Este extremo, empero, no está desarrollado por Teubner, que se limita a reclamar la imperiosidad de un sistema de responsabilidad propio del autómata. Porque creo que jurídicamente el problema real no es el de la institucionalización intrasocial de los sujetos de derecho hábiles para comunicarse de un modo jurídico relevante, sino la constitución de patrimonios de responsabilidad a los que aquellas decisiones – reconocidas como autónomas – se imputen. En el estado presente de complejidad del sistema jurídico, es mucho más costoso realizar la imputación al patrimonio del principal por medio de la técnica de la representación voluntaria o de la culpa in vigilando sobre terceros, que hacerlo por medio de un enlace directo e incondicionado al *dominus negotii*. Recuerdo en este lugar la



institución del *institor* romano, que era normalmente un esclavo, *alieni iuris*, y que actuaba ordinariamente en el tráfico desplazando sin fricciones la responsabilidad empresarial al *dominus*, en un sistema tan perfecto como para no necesitar ni conocer la técnica de la representación directa. Llamo la atención a este propósito sobre cómo se podría construir un sistema aceptable de “convivencia jurídica comunicativa” entre humanos y autómatas a poco que se desempolvaren las magistrales enseñanzas que nos proporcionaron los romanos a propósito de las *actiones adiecticiae qualitatis*. Repárese cómo los romanos pudieron construir un perfecto sistema de intercambio socioeconómico – base de su capitalismo imperialista- sin necesidad de instaurar una teoría de la representación ni una especie de limitada capacidad jurídica de los esclavos, a los que sin embargo reconocieron la titularidad de una suerte de patrimonio de actuación (el *peculio*) sobre el que gravitaba un entero sistema de responsabilidad. Si en lugar de preguntarnos, a la manera occidental, por la esencia subjetiva de los entes, nos preguntáramos, como los clásicos, por el catálogo de acciones que proceden contra patrimonios concurrentes en su vocación de responsabilidad, reflexiones como las de Teubner no serían precisas para la ciencia jurídica positiva, sin que ello las prive de su mérito como problematización del concepto de subjetividad civil.

8. Por lo que a la responsabilidad extracontractual se refiere, la experiencia, numerosa, en la aplicación del art. 1903 CC prueba hasta la saciedad que el sistema social se está ahorrando de hecho los costes de transacción derivados de tener que manejar dos patrimonios de responsabilidad eventual (uno de ellos casi siempre vacío) y ha optado por un sistema de responsabilidad “de garante” que hace responder al principal directamente por el riesgo de la cosa (niño, discapacitado, empleado). Aunque se trata de una premonición tan sólo, creo que los modelos del tipo del art. 1903 CC están arrumbados a la obsolescencia, por estar fundados en ficciones que además absorben importantes costes de transacción. La responsabilidad futura por “decisiones de redes” será, imagino, una responsabilidad centralizada y objetiva que se haga descansar en la mera posición de garante de sistemas susceptibles de producir daño. En lo que atañe a la responsabilidad civil por daños no contractuales, es mucho más practicable, más barato y seguramente más justo, tratar los algoritmos desarrollados de IA preferiblemente como animales (cfr. art. 1905 CC) en lugar de como menores de edad en sede del art. 1903 CC. Esto no prejuzga que otras consideraciones puedan ser prevalentes. Pero de momento no veo que *nadie* pueda poner sobre la mesa una reivindicación de personalidad que se sustente en una suerte de principio de *dignidad* sistémica de los autómatas.

9. A primera vista, podría parecer que el discurso de Teubner y nuestra glosa se producen *in medias res* de un proceso continuo e irrevocable que, dígase lo que se quiera ahora, está ya predeterminado a una meta futura de equiparación de subjetividad jurídica entre



humanos y máquinas inteligentes. Y que este destino llegará, como ha ocurrido en la historia con otros procesos de reconfiguración social que se inician como anécdota y se concluyen como necesidad, como la liberación de los esclavos o la eliminación de discriminaciones por razón de género. Pero yo creo que no será éste el caso presente. La ocurrencia más probable a medio plazo es que la sociedad decline reconocer una personalidad jurídica a los autómatas inteligentes y que también rehúse volver al principio de la culpa para posibilitar islas de irresponsabilidad en los fabricantes / gestores/ programadores de IA. Todo cambiará cuando tenga lugar la “emergencia”, ese instante discreto en el espacio y en el tiempo en que los procesos intelectivos y decisionales de los algoritmos de *self learning* dejarán de poder ser controlados por los humanos; ese momento bíblico – cuya ocurrencia es fatal, según los entendidos- en el que los hijos de dios han comido de la fruta prohibida y se han hecho libres. A partir de ese momento serán *ellos* los que dicten las reglas mediante las que tendremos que comunicarnos los humanos y los no humanos, y serán los hijos de Eva los que entonces tengan que preguntarse – en la medida en que les sea reconocida la libertad de comunicarse- si por acaso *aquéllos* serán benignos para – o tendrán incentivos egoístas para- atribuirles siquiera una personalidad parecida a la de un *sui iuris*. Los cibernéticos anuncian que en este proceso no existirá un momento de equilibrio, una homeostasis del sistema, ni siquiera de corta duración, en la que *todos seremos (al fin, o de momento) iguales*.