

# ESTADO DEL ARTE EN INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS

(Extracto de la propuesta de creación de la Red de Inteligencia Artificial y Ciencia de Datos,  
septiembre de 2019)

Juan Humberto Sossa Azuela

Alejandro Peña Ayala

## 1. Introducción

La *revolución industrial* (RI) es un referente para caracterizar la evolución de la sociedad a través de diversos factores como son: actividad humana, conocimiento, capital, energía, materia prima, medios de manufactura y oferta de productos y servicios que contribuyen a generar riqueza y progreso tal y como se aprecia durante cuatro generaciones [1]. La *primera. RI*, ocurrida entre 1760 y 1830, se distinguió por la mecanización sustentada en la energía de vapor producida por agua y carbón como insumos [2]. En tanto que la *segunda. RI*, sucedida entre 1870 y 1914, se caracterizó por la fabricación en serie de bienes, las líneas de ensamblado y el uso de energía eléctrica [3]. En la mitad del siglo XX, el surgimiento de las computadoras y la automatización de los procesos de manufactura mediante el uso de energías renovables dieron vida a la *tercera RI*, también denominada científico–tecnológica [4]. Mientras que en el Foro Económico Mundial 2016 su fundador Klaus Schwab anunció el surgimiento de la *cuarta. RI*, en la cual *fábricas inteligentes* son capaces de auto–controlarse en cada etapa de la cadena de valor [5], donde energías limpias alimentan sistemas ciberfísicos encargados de controlar procesos físicos, recrear ambientes virtuales y tomar decisiones autónomas gracias al uso de *inteligencia artificial* (IA), *ciencia de datos*<sup>2</sup> (CD), Internet de las cosas, nanotecnología, robótica, biotecnología y la computación en la nube entre otros campos.

En el marco delineado por la presente *cuarta RI*, la gente y todos los sectores de la sociedad están inmersos en un universo conformado por datos, mensajes, transacciones y objetos multimedia que se expande aceleradamente como resultado del poder de cómputo de los ordenadores, la mecanización y automatización de una creciente variedad de aplicaciones, el uso generalizado del Internet y dispositivos inteligentes móviles, y la diversificación de las redes sociales [6]. Esta *explosión de contenido digital* genera estímulos sensoriales que inducen un estado de alerta en las personas, elevan su carga cognitiva, demandan desarrollar sus capacidades mentales y exigen regular sus emociones en aras de que los individuos sean capaces de percibir, comprender, analizar y responder con atinencia ante tal demanda para con ello desempeñarse hábilmente en un entorno dinámico y heterogéneo [7]. Por su parte, las organizaciones públicas y privadas afrontan un entorno globalizado, caracterizado por una compleja, volátil y estocástica red de actores, fenómenos e interdependencias que les exige realizar con eficiencia la *gestión de datos masivos, información y conocimiento*. Toda vez que la organización contemporánea requiere acceder flujos de datos voluminosos y continuos, transformar tales repositorios en información significativa, examinar dichos acervos para descubrir conocimiento útil que les faculte a tomar decisiones acertadas, elevar la productividad, mejorar la calidad de productos y servicios, e incrementar el grado de satisfacción del consumidor para con ello competir exitosamente en su entorno [8].

Ante un escenario tan demandante para el individuo, la sociedad y las organizaciones, la IA y la CD cobran especial relevancia como instrumentos de asistencia, capacidad y valor. En cuanto a la IA, este término fue acuñado por John McCarthy en 1955 con el propósito de organizar (junto con Claude Shannon, Marvin Minsky y Nathaniel Rochester) el Proyecto de Investigación de Verano 1956 en Dartmouth sobre IA, en el cual la hipótesis fue: “Cada atributo del aprendizaje o cualquier otra característica que distingue la inteligencia es posible, en principio, ser descrita con tal precisión que una máquina puede ser construida para simularla” [9, 10]. Dada la ambiciosa proposición y los novedosos conceptos que entrañaron dicha iniciativa y ante la invención de las computadoras, súbitamente emergieron diversas líneas de investigación, tales como: prueba de teoremas [11], aprendizaje automático de máquina [12], sistemas expertos [13], sistemas tutores inteligentes [14], reconocimiento de patrones [15], visión artificial [16] y procesamiento de lenguaje natural [17]. Ante la diversidad de intereses, la IA paulatinamente empezó asociarse con otras ramas de estudio, y en cierto modo a integrarlas bajo su mismo término, como por ejemplo: redes neuronales

artificiales [18, 19], cibernética [20], computación evolutiva (p. ej. estrategias evolutivas [21], programación evolutiva [22], algoritmos genéticos [23]), lingüística [24], ciencias cognitivas [25], inteligencia computacional [26], el cómputo suave [27] y la robótica [28]. Es tal el alcance y amplitud de la visión de la IA, que de acuerdo con Elaine Rich: “La IA es el estudio que busca hacer que las computadoras realicen actividades intelectuales que por el momento la gente las hace mejor” [29]. Además, conforme las aplicaciones computacionales gradualmente logran un desempeño eficiente, entonces ellas dejan de ser competencia de la IA, pues a ésta solo le incumbe aquello que aún no ha sido implementado satisfactoriamente.

Por lo que respecta a la CD, su antecedente proviene del término *análisis de datos*, cuyo concepto, definido por John Wilder Tukey en 1962, establece: “Son las formas de planificar la recopilación de datos, los procedimientos basados en las estadísticas matemáticas que se aplican para examinar los datos, los resultados y las técnicas para interpretarlos de manera fácil, precisa y acertada” [30]. Posteriormente en 1974, Peter Naur propone el término CD para afirmar: “Es la ciencia del tratamiento de datos, una vez que ellos han sido establecidos, mientras que la relación entre los datos y lo que ellos representan corre a cargo de otros campos y ciencias” [31]. Más tarde en 1976, la Federación Internacional de Sociedades de Clasificación intitula su conferencia “CD, Clasificación y Métodos Relacionados”, y un año después la Asociación Internacional para la Estadística Computacional se funda para relacionar la metodología estadística tradicional, la tecnología computacional y el aporte de expertos para convertir datos en información y conocimiento, mientras que durante en curso en la Universidad de Michigan Jeff Wu propone renombrar a los estadísticos como *científicos de datos* [32]. En resumen, para Foster Provost y Tom Fawcett la CD es: “Un conjunto de principios que soportan la extracción de información y guían la generación de conocimiento inmersos en acervos de datos masivos” [33]. Si bien el fundamento esencial de la CD descansa en la estadística, el término se emplea actualmente como un aglutinador de diversas líneas de trabajo como por ejemplo: procesamiento de datos electrónico [34], sistemas de información mecanizados [35], bases de datos [36], proceso de datos en línea [37], sistemas distribuidos [38], almacenes de datos [39], inteligencia de negocios [40], proceso analítico en línea [41], descubrimiento de conocimiento en bases de datos [42], datos masivos [43], cómputo en la nube [44], minería de datos [45], analíticas [46] y visualización de datos [47].

Una vez enunciado el contexto que la *cuarta RI* recrea en la actividad productiva de la sociedad contemporánea, donde la gente afronta la *explosión de contenido digital* y las organizaciones son exigidas a *gestionar datos masivos, información y conocimiento*, se ha considerado a la IA y la CD como disciplinas en vías de maduración capaces de ofrecer constructos teóricos, metodológicos y aplicativos para empoderar individuos y organizaciones a encarar tales retos cognitivos y logísticos, además de habilitar los medios para que ellos desarrollen los sistemas ciberfísicos que dan vida a las *fábricas inteligentes*. En este escenario, ilustrado en la Fig. 1, la CD regula el ciclo compuesto por la explotación de acervos de datos, su transformación en información, y el análisis de ésta para generar conocimiento como soporte a la toma de decisiones. Mientras que la IA procura adquirir y representar conocimiento para buscar e inferir las soluciones a problemas complejos que demandan la emulación de inteligencia humana, mismas que son implementadas como aplicaciones que interactúan con su entorno y reguladas para elevar su eficiencia. De este modo, investigación, academia e industria vinculadas con la IA y CD proponen una plataforma inteligente y basada en datos que ofrece un marco de trabajo e infraestructura útil para sistematizar el examen de todo aquello que caracteriza la labor humana, proveer significado informativo dado un contexto, descubrir elementos racionales que abonan a cobrar conciencia de un caso determinado, y formular respuestas a cuestiones que demandan un alto grado de saber, pericia y desempeño.

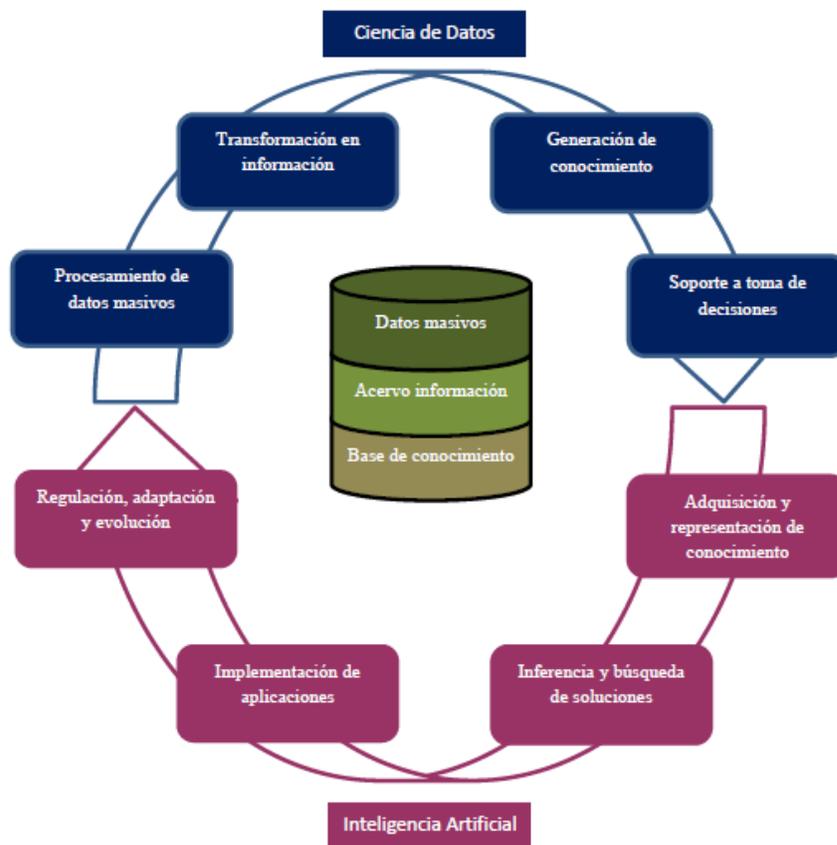


Fig. 1 Ciclo colaborativo IA – CD compuesto por las etapas de trabajo que las componen.

Como resultado del proceso transformador encausado por las dos primeras RIs, surgieron las primeras potencias mundiales que beneficiaron a sus sociedades como consecuencia de su liderazgo educativo, científico, tecnológico, empresarial y comercial, los cuales elevaron el nivel de la economía y bienestar social de sus pueblos, como el británico y el americano. En contraparte, la mayoría de los países no invirtieron debidamente en dichos campos, por lo que comenzaron a rezagarse en la medida que apareció y creció una brecha en todos esos ámbitos que los ha puesto en desventaja y dependencia frente las naciones desarrolladas.

En la presente sección se ha enunciado el avance de la sociedad mundial a la luz de las *cuatro RIs* y reconocido su impacto en las personas y organizaciones, introduciendo la IA y CD como disciplinas que contribuyen a encausar la transformación y hacer frente a sus demandas.

## 2. Campo de la Inteligencia Artificial

Una vez que el origen, los conceptos, las líneas de trabajo y ramas de la IA han sido expuestos en la Introducción, en esta sección la perspectiva esencial de la IA se complementa por medio de un perfil y se enriquece mediante la exposición de las líneas de generación y aplicación de conocimiento (LGAC), las cuales representan los temas que inspiran la investigación y el desarrollo tecnológico orientado a contribuir al conocimiento e impacto de la IA.

### 2.1 Perfil de la inteligencia artificial

Toda vez que la *inteligencia* está relacionada con la psique (entendida esta como los procesos conscientes e inconscientes propios de la mente), el referente clásico para determinar si un ente es inteligente se relaciona con la psicometría (disciplina que mide sistemáticamente ciertas propiedades psicológicas a través de instrumentos). Por tanto, la IA no está exenta de tal pretensión, siendo el “Test de Turing” su principal patrón de evaluación [48]. A partir de tal referente, la visión de la IA es concebir y crear entidades artificiales

capaces de aprobar el “Total Turing Test” [49], en el cual un agente no solo debe exhibir un desempeño lingüístico similar al humano, sino también una actuación sensorial y motora idéntica [50].

Además del reto que representa generar sistemas inteligentes que satisfagan criterios como los psicométricos, otro corresponde aquellos que ostentan la propiedad *artificial*. Es decir que a partir de los enfoques *creacionista científico*<sup>1</sup> y *diseño inteligente*<sup>2</sup>, la IA en principio, es también un *sistema artificial*<sup>3</sup> concebido, creado y desarrollado por humanos, que idealmente incluye capacidades que le permiten transitar por un ciclo de *vida artificial*<sup>4</sup> al considerar complejos paradigmas, tales como: auto-organización<sup>5</sup>, sistemas evolutivos<sup>6</sup>, autopoiesis<sup>7</sup>, teoría general de sistema<sup>8</sup>, homeostasis<sup>9</sup>, auto-reproducción<sup>10</sup>, autómatas que se auto-reproducen<sup>11</sup>, auto-regulación<sup>12</sup>, y teoría de la información<sup>13</sup>.

Amén de las exigencias científico-tecnológicas que implica recrear la naturaleza de los dos términos que componen el nombre de la IA, existe un tercer factor a considerar en la gestación de aplicaciones y sistemas de la IA, el alcance [65]. Es decir, la investigación, avances y productos generadas por la IA (desde los esfuerzos por demostrar teoremas y resolver problemas en los albores de la IA [66], hasta el éxito logrado por AlphaGo<sup>14</sup> y Eugene Goostman<sup>15</sup> [68]) se caracterizan por estar orientados a desarrollar aplicaciones que exhiben un alto grado de pericia en el ejercicio de una función compleja, al grado de superar al desempeño humano. Sin embargo, pese a lo impresionante del logro, estas aplicaciones tienden a ser consideradas como “IA débil” [69], toda vez que en contraste con la inteligencia humana, ésta es capaz de encarar con éxito una amplia variedad de funciones complejas. Por lo tanto, el reto es crear “IA fuerte” [70] o “IA general” [71] capaz de realizar eficazmente diversas funciones complejas como el humano acostumbra realizar en su cotidianidad.

Con respecto a la relación que la IA guarda con la *computación*, tradicionalmente se considera que la primera es una rama de la segunda [72]. Sin embargo, es pertinente enunciar qué si bien la IA reconoce la influencia ejercida por la computación en su concepción, evolución y futuro desarrollo, en realidad se trata de un par de campos de estudio, desarrollo y aplicación con alto grado de afinidad conforme a las siguientes razones. En esencia la IA es un área de conocimiento interdisciplinaria [73], donde convergen diversas ciencias como: neurociencia, psicología, cognición, computación, informática, biología, lingüística, filosofía,

---

<sup>1</sup> Examina todos los aspectos pertinentes del tema del origen exclusivamente desde bases científicas y sin hacer referencia a la teología [51].

<sup>2</sup> Afirma que ciertas propiedades del universo y de los seres vivos se explican mejor por una causa inteligente, en vez de un proceso no dirigido como el de la selección natural [52]

<sup>3</sup> Es un sistema creado por el humano y adaptado para alcanzar objetivos y propósitos humanos [53]

<sup>4</sup> Constructo computacional orientado a sintetizar propiedades y fenómenos normalmente asociados con los sistemas vivos naturales [54]

<sup>5</sup> Revela la habilidad de un sistema para exhibir patrones espacio-temporales ordenados solamente como el resultado de las interacciones entre sus componentes [55]

<sup>6</sup> Estudia y caracteriza los mecanismos y leyes de la evolución biológica [56]

<sup>7</sup> Representa aquello que ocurre en la dinámica de la autonomía propia de los sistemas vivos [57]

<sup>8</sup> Marco teórico, matemático y filosófico de la síntesis con orientación holística, ecológica e integrativa, que estudia la naturaleza organizada y activa de los organismos vivos, considerados como sistemas abiertos [58]

<sup>9</sup> Capacidad del organismo para mantener condiciones internas de estabilidad ante cambios del entorno [59]

<sup>10</sup> Sistemas cuyas condiciones de identidad al paso del tiempo incluyen su propia reproducción continua [60]

<sup>11</sup> Aborda el problema de hacer máquinas capaces de reproducirse ellas mismas mecanísticamente como una forma que revela cuestiones de la biología y las capacidades y limitaciones de las máquinas [61, 62]

<sup>12</sup> Habilidad del individuo de controlar estados internos y respuestas a pensamientos, emociones, atención y desempeño como parte esencial del desarrollo de la personalidad y conducta [63]

<sup>13</sup> Analiza la cuantificación, almacenamiento y comunicación de información, estimando constructivamente distribuciones probabilísticas con conocimiento parcial para inferir la estimación máxima de entropía [64]

<sup>14</sup> Aplicación de IA hecha por Google DeepMind para jugar Go, que ha vencido a los mejores jugadores [67]

<sup>15</sup> Aplicación de IA que representó a un adolescente sometido a prueba en la Royal Society of London, habiendo hecho creer al 33 del jurado que era humana y con ello aprobado el “Test de Turing” [68]

sociales, ingeniería de sistemas, control y de manera especial las matemáticas [74]. Así mismo, la IA emplea como principal instrumento de trabajo los recursos de hardware, software y comunicación que la computación y las telecomunicaciones [75] le brindan como un medio para crear e implementar sus propios sistemas, los cuales oscilan desde el uso de supercomputadores [76] y ambientes distribuidos [77], hasta los dispositivos móviles [78].

Además de tales distingos, los atributos que caracterizan el trabajo realizado por la IA (expuestos en todas las secciones del documento, donde destacan: inteligencia, artificial y alcance) son otro factor que resalta la identidad de la IA ante la computación. Aunado a ello, hay también otras características que ostentan los problemas y soluciones que son abordadas por la IA que son dignas de ser mencionadas, las cuales contrastan contra los respectivos atributos de los casos que por lo general afronta la computación, tales como:

- *Ignorancia:* La IA carece del conocimiento que garantice una solución al problema en turno, pues al contar con él entonces se trataría de un caso tradicional de la computación.
- *Factibilidad:* Los problemas abordados por la IA pueden carecer de solución, tener una o varias; en contraste con los abordados por la computación que sola tiene una respuesta.
- *Representación:* La generación de la solución es ilustrada por la IA como un espacio de estados (inicial(es), intermedio(s), final(s)) que están entrelazados por medio de una secuencia de operaciones, en contraste con el flujo de trabajo clásico de la computación.
- *Instrumentación:* En la confección de una solución, la IA conjuga un marco teórico compuesto por búsqueda, conocimiento, razonamiento y aprendizaje; en contraposición al dúo algoritmo y datos que compone las propuestas de la computación.
- *Saber:* La IA exige adquirir y representar conocimiento especializado del caso, mismo que usa para caracterizar los estados, operaciones y soluciones; mientras que la computación manipula datos como insumo que transforma en información.
- *Exploración:* La IA desarrolla una solución a través de la *búsqueda* en un espacio de estados que paulatinamente va generando (para lo cual usa diversos criterios y estrategias, como: ciega, heurística, a profundidad, por amplitud, hacia adelante, hacia atrás...) como resultado de aplicar una secuencia de operaciones a los estados (inicial(es), intermedio(s) o final(es)) por lo que se produce un grafo compuesto por multitud de estados y rutas que los ligan; en contraparte a soluciones representadas por una ruta específica.
- *Relatividad:* Cuando el caso ostenta varias soluciones, la IA usa criterios para evaluarlas y reconocer la ideal u óptima de las buenas; en contrapunto a las soluciones únicas.
- *Estimación:* La IA usa diversos mecanismos de razonamiento para calcular la transición entre los estados (inicial(es), intermedio(s) o final(es)) que integran el espacio de estados, tales como: inductivo, deductivo, abductivo, probabilístico, no monótonico, temporal difuso, no determinístico...; en contraste al proceso cuantitativo de la computación.
- *Adaptación:* La IA concibe soluciones con capacidad para aprender conocimiento del problema, depurar estrategias de búsqueda en el espacio de estados y refinar procesos de razonamiento en aras de evolucionar gradualmente hasta lograr soluciones satisfactorias; en cambio la computación depura manualmente la solución hasta alcanzar la óptima.
- *Pluralidad:* La solución al problema que aborda la IA es de diversa naturaleza, como por ejemplo: generar el estado final, arribar al estado inicial, una secuencia de operaciones que transitan el espacio de estados (desde el inicial al final o viceversa), los criterios que evalúan el estado que marca la solución al problema, o bien los factores considerados para desarrollar la secuencia; en contraparte con el afán de calcular el estado final.

- *Aproximabilidad*: La IA generalmente privilegia la aproximación a la solución ideal y las buenas soluciones en disparidad a la solución óptima procurada por la computación.
- *Incertidumbre*: En el desarrollo de la solución la IA no tiene certeza absoluta de la naturaleza de los estados que genera, el efecto que produce una operación sobre un estado dado (inicial, intermedio o final), ni la conveniencia de la secuencia de operaciones; en contrariedad a la certeza que gozan las aplicaciones de la computación.
- *Heterogeneidad*: El conocimiento que manipula la IA se caracteriza por ser cualitativo, subjetivo, difuso, incompleto, impreciso, dinámico, inestable, complejo, contradictorio, estocástico, y aproximado entre otros rasgos; en contraste con las cualidades antónimas que de forma respectiva ostenta el conocimiento que manipula la computación.
- *Aproximación*: La IA procura desarrollar una solución a través del “ensayo y error” en forma manual o (semi) automática con el objeto de generar soluciones progresivamente mejores; en oposición a la computación que crea una solución óptima en primer instancia.

## 2.2 Líneas de generación y aplicación de conocimiento de la inteligencia artificial

Con respecto a las LGAC que en materia de IA la *Red* propone desarrollar, se encuentran las siguientes:

- *Básica*: Procura el avance teórico y conceptual de frontera que contribuye al desarrollo de la IA mediante la propuesta de: teorías, leyes, hipótesis, marcos de trabajo, modelos, metodologías e instrumentos que consolidan sus fundamentos [79].
- *Formal*: Insta a desarrollar los constructos matemáticos y filosóficos que sustenten el diseño de aportes teóricos y aplicaciones prácticas de la IA [80].
- *Especializada*: Incursiona en tendencias enmarcadas por paradigmas como: IA fuerte, IA débil, súper-IA, IA avanzada, IA descentralizada. IA 2.0... [81].
- *Vida artificial*: Concepción de organismos vivos, artificiales e inteligentes que emulen y caractericen cualidades, comportamientos y ciclo de vida semejante al natural [82].
- *Mente artificial*: Recrea propiedades, funciones y cualidades características de la mente natural como: conciencia, pensamiento, emociones, sentido común... [83].
- *Cognición*: Representa atributos y capacidades cognitivas naturales, tales como: metacognición, aprendizaje, memoria, razonamiento, modelo del cerebro... [84].
- *Conocimiento*: Innova esquemas para la representación, semántica y explotación del saber acumulado y distribuido en el ámbito digital como ontologías, Web 3.0... [85].
- *Búsqueda*: Perfecciona procesos y estrategias para explorar soluciones en espacios de estados a través de: heurísticas, selección atributos, ambientes distribuidos... [86].
- *Razonamiento*: Formula procedimientos para producir conocimiento racional del tipo: difuso, lógico, Bayesiano, sentido común, analógico, baso en casos, neural... [87].
- *Aprendizaje*: Explora métodos para el aprendizaje automático de máquina, profundo, (no) supervisado, reforzado, enjambre, estadístico, conexionista... [88].
- *Computacional*: Conjuga esfuerzos con líneas de computación afines, tales como: suave, inteligencia, evolutiva, neural, paralela, distribuida, en la nube... [89].
- *Funciones*: Desarrollo de funcionalidades vinculadas con: percepción, lenguaje, visión, reconocimiento de patrones, interfaces, imagen, agentes... [90].
- *Maquinaria*: Integración de facultades inteligentes, adaptivas y evolutivas en robots, humanoides, circuitos, equipos y cobots para lograr desempeño eficiente [91].
- *Acervos*: Vinculación con bases de datos, datos masivos, sistemas expertos y basados en conocimiento, multimedia, minería, analíticas, toma de decisiones... [92].

- *Ciencias*: Apoyo al avance y aplicación de disciplinas: física, biología, neurología, química, astronomía, geología, ingeniería, nano, información... [93].
- *Actividad productiva*: Soporte para elevar la eficiencia en la administración, banca, agricultura, finanzas, fiscal, contable, mercadotecnia, turismo, negocios... [94].
- *Industria*: Como medio para mejorar productividad y calidad en industrias de energía, manufactura, transformación, aeronáutica, espacial, automotriz, servicios... [95].
- *Comunicación*: Control del uso: telecomunicaciones, redes sociales, cómputo móvil, ubicuo y pervasivo, domótica, chatbot, Internet de la cosas, telefonía celular... [96].
- *Seguridad*: En el aras de reforzar protección en ámbitos como: defensa, ciber-ataques, pública, privada, identidad, intrusión, virus, secuestro físico y virtual... [97].
- *Asistentes*: Oferta de apoyo al usuario: sistemas de recomendación, compañero personal inteligente, adaptación y personalización, compañero virtual... [98].
- *Gobierno*: Soporte para el ejercicio de la administración pública, labor legislativa e impartición de justicia en contextos locales, nacionales e internacionales [99].
- *Educación*: Adaptación y personalización de la enseñanza y aprendizaje a través de: tutores, modelado del usuario, soporte y evaluación centrada en el estudiante... [100].
- *Salud*: Auxilio en la preservación de la salud y extensión de capacidades humanas: cyborgs, órganos y prótesis artificiales, nanobots, mapeo cerebral... [101].
- *Gobernanza*: Concierno a los aspectos de gestión, política, penetración social, ética, privacidad, vigilancia, protección, marco legal... [102].
- *Bienes*: Diseño, prueba, producción y/o operación de: vehículos autónomos, ropa, aparatos electrodomésticos, asistente hogar, vigilancia, televisores... [103].
- *Ambientes*: Preservación del medio ambiente, diseño y recreación de: ciudades, fábricas, edificios, casas, laboratorios, aulas, salas... inteligentes [104].
- *Movilidad*: Asistencia en la planificación y regulación de urbanismo, transporte, tráfico, rutas, señalización, sanción, manejo de siniestros, abasto energía... [105].
- *Artes, cultura y entretenimiento*: Revolución del quehacer artístico, cultural y del entretenimiento, tales como literatura, exposiciones, juegos... [106].
- *Instrumentos*: Creación de métodos y programas de cómputo para uso en aplicaciones de IA y que son objeto de propiedad intelectual ante IMPI e INDAUTOR [107].

### 3 Campo de la Ciencia de Datos

De acuerdo con los antecedentes, conceptos básicos y líneas de trabajo que distinguen de la CD manifestados en la Introducción, así como las etapas a su cargo en el ciclo colaborativo con la IA, en la presente sección se complementa un panorama de la CD mediante un perfil que describe sus atributos y elementos, que es acompañado por una serie de instancias que componen las LGAC.

#### 3.1 Perfil de la ciencia de datos

El término *ciencia* proviene del latín “scientia” que significa *conocimiento*, el cual se deriva de un sistema racional de saberes adquiridos del estudio de fenómenos naturales, sociales y artificiales, por tanto “ciencia es una estructura construida sobre hechos” [108]. Por su parte, *datos* significa “una serie de signos que constituyen medios de comunicación para expresar la naturaleza y actividad humana” [109]. Los datos se organizan en forma estructurada<sup>16</sup>, semiestructurada<sup>17</sup>, o no estructurada<sup>18</sup> [110]. Entonces, dado que los hechos se pueden expresar como datos, la CD es: “una estructura construida sobre hechos caracterizados como signos que facilitan la comunicación de fenómenos naturales, sociales y artificiales”.

Asimismo, al conjunto de signos que puede recrear un significado dado un contexto se le dice *información* [111]. La información se revela como la ocurrencia del proceso de comunicación entre el emisor y el receptor del mensaje. Es decir, todo aquello que los datos comunican es *información* [112]. Cuando los datos expresan aquello que ocurre en la naturaleza o actividad humana es, información, por lo que su utilidad es la información que brindan [113]. Entonces, CD es: “una estructura fundada en hechos revelados como una serie de signos que dan sentido al proceso de comunicación de fenómenos naturales, sociales y artificiales”.

A ese respecto, *conocimiento* es el resultado de la selección de significado de un sistema psíquico o social de la naturaleza o actividad humana derivado de la comunicación provista por la información [114]. Es decir, conocimiento representa un estado cognitivo o afectivo que revela significado y entendimiento. Por lo que conocimiento es un marco de referencia personal o cognitivo que hace posible usar información. Con relación a datos e información, el conocimiento es más subjetivo e intangible, ya que es el referente que el individuo toma de ellos para incorporarlo a su sistema de creencias, valores, procesos y acciones. El conocimiento exige datos, información, el significado de los signos (el cual es determinado por la posición del signo en un espacio de signos) y el contenido de los signos (el cual es denotado por la posición del elemento –denotado por el signo– en el espacio de elementos [115]. Por lo consiguiente, la CD es: “una estructura fundada sobre hechos revelados como una serie de signos que otorgan sentido, significado y contenido de un sistema psíquico o social del proceso de comunicación de fenómenos naturales, sociales y artificiales”.

En aras de analizar la trilogía DIC: datos (D), información (I) y conocimiento (C) se aplican las siguientes definiciones: *D* son estímulos sensoriales que son percibidos por los sentidos; *I* es el conjunto de datos procesados en una forma que es significativa para el receptor; *C* es aquello que es aprendido, entendido y evaluado por el individuo [116]. Así mismo, se aplica un enfoque de las ciencias de la información [117], el cual es caracterizado por elementos propositivos, centrados en el humano, cognitivos y no metafísicos, a partir de los cuales se definen dos dominios, externo (en el entorno del individuo) e interno (en la mente de la persona). Con base en dichos componentes se diseñan cinco modelos de DIC [118]: 1) universal: D, I; subjetivo: C; 2) universal: D; subjetivo: I, C; 3) universal: D, I, C; subjetivo: I, C; 4) universal: D, I; subjetivo: D, I, C; 5) universal: D, I, C; subjetivo: D, I, C [119].

A partir de dichos constructos, la CD forma parte de un flujo de trabajo compuesto por: 1) la ingeniería de datos y el proceso de datos masivos; 2) el proceso de CD; 3) la toma de decisiones (manual y automática)

---

<sup>16</sup> Ceñidos a un formato específico como una base de datos

<sup>17</sup> No aplican un formato estricto como ocurre con un documento eXtended Markup Language

<sup>18</sup> Codificados de manera heterogénea o con un formato desconocido, como son los objetos multimedia

dirigida por datos [120]. La CD es una disciplina que integra la estadística, bases de datos y minería de datos [121]. A ese respecto, un diagrama de Venn clásico de la CD consta de tres campos: 1) conocimiento matemático y estadístico; 2) Destreza sustantiva sobre el dominio de la aplicación de datos; 3) habilidades de intrusión. Donde la intersección de 1 y 2 implica los métodos de investigación tradicional, la conjunción de 2 y 3 corresponden a una zona de riesgo, en la mezcla entre 1, 3 aparece el aprendizaje automático [122]. De acuerdo con [123], el perfil de CD se amplía con las ciencias de la computación, comunicación y habilidades de presentación, y visualización, además de incluir un proceso compuesto por: 1) recopilación de datos crudos y heterogéneos; 2) proceso de datos; 3) limpieza de datos; 4) análisis exploratorio de datos; 5) aplicación de algoritmos de aprendizaje automático y generación de modelos estadísticos; 6) visualización, reportes y comunicación para la toma de decisiones; 7) construcción del producto de datos. En este contexto, el rol que juega el *científico de datos* es fundamental para realizar con habilidad las etapas del proceso con el fin de elegir los métodos y herramientas apropiados.

### 3.2 Líneas de generación y aplicación de conocimiento de la ciencia de datos

Una vez que ha sido expuesto el concepto para los términos ciencia, datos y CD, éste último con tres acepciones correspondientes a la relación DIC, ha quedado claro que la CD es un esquema recreado por hechos que son representados por datos (expresados como signos), los cuales son transformados en información, a partir de la cual se genera conocimiento acerca de los fenómenos naturales, sociales y artificiales. Siendo este último el insumo para la toma de decisiones manuales y automatizadas, así como el vínculo con la IA quien lo aprende y representa para usarlo en procesos de búsqueda e inferencia de soluciones, mismas que aplica y regula en la forma descrita e ilustrada en la Introducción y Fig. 1. En consecuencia, resulta lógico que la variedad de LGAC propuestas para la CD coincidan en su denominación, más no en su naturaleza, con las propias de la IA. Por lo tanto, la *Red* se orienta a desarrollar las siguientes LGAC:

- *Básica*: Contribuye al progreso teórico y conceptual del estado del arte de la CD al proponer: teorías, principios, modelos, marcos de trabajo y metodologías [124].
- *Formal*: Procura consolidar los fundamentos matemáticos, estadísticos y algorítmicos de la CD que otorgan veracidad y certidumbre a sus aplicaciones y resultados [125].
- *Especializada*: Busca iniciar o participar en tendencias relacionadas con: sabiduría, pensamiento de CD, cerebro de datos, complejidad, CD grande, gran escala... [126].
- *Inteligencia*: Participa en líneas que distingan diversos tipos de inteligencia como: integral, conducta, dominio, organizacional, social, red, datos, ambiente... [127].
- *Conocimiento*: Cooperar en varios aspectos vinculadas con el conocimiento, a saber: epistemología, descubrimiento, semántica, bases de conocimiento... [128].
- *Aprendizaje*: Diseño de métodos de búsqueda hacia atrás y adelante, aprendizaje (no y semi) supervisado, refuerzo, simbolista, conexionista, evolutivo, análogo... [129].
- *Orientado a DIC*: Une esfuerzos con líneas afines, tales como: minería de datos y de texto, analíticas, inteligencia de negocios, proceso analítico en línea... [130].
- *Repositorios*: Explotación de bases de datos, almacenes de datos, datos masivos, datos distribuidos en la nube, objetos multimedia, datos no estructurados... [131].
- *Ciencias*: Tratamiento de DIC relacionado con diversas disciplinas: estadística, ciencias, biología, física, sociales, genómicas, climatología, de la tierra... [132].
- *Actividad productiva*: Como soporte para las actividades relacionadas con: la toma de decisiones, predicción, negocios, turismo, administración, detección fraudes... [133].
- *Industria*: En asistencia a la planeación y control de las actividades de energía, cadena de suministro, manufactura, servicios, transformación, servicios... [134].

- *Comunicación*: Rastreo, regulación e interpretación del uso de comunicación móvil, de redes sociales, streaming de datos, sonido y video, Internet de las cosas... [135].
- *Seguridad*: Análisis proactivo de amenazas de intrusión, robo de datos, quebranto de privacidad, ciberataques, copia de identidad, secuestro virtual, vigilancia... [136].
- *Asistentes*: Respaldo al usuario en sus actividades: navegación en Internet, compra electrónica, sistemas de recomendación, interacción, adaptación, interfaces... [137].
- *Gobierno*: Como medio para mejorar la práctica de la administración pública, función legislativa y aplicación de justicia a nivel local, nacional e internacional [138].
- *Educación*: Uso de funciones proactivas y reactivas para centrar la educación en el estudiante, elevar aprendizaje, reducir deserción, abatir prácticas ilegales... [139].
- *Salud*: Contribución a la práctica médica, salud pública, prevención y combate de epidemias, tratamiento al paciente, eficacia de medicamentos y sus efectos... [140].
- *Gobernanza*: En aras de regular el acceso, organización, explotación y diseminación de DIC la CD provee “datos duros” para formular códigos, reglas éticas... [141].
- *Bienes*: Provisión de asistencia en la concepción, diseño, examen, fabricación, uso y mantenimiento de vehículos autónomos, maquinaria, equipo... [142].
- *Ambientes*: Auxilio en el diseño, desarrollo y preservación del medio ambiente, y del funcionamiento de ciudades, fábricas, inmuebles, espacios inteligentes [143].
- *Movilidad*: Apoyo en la planeación y control del urbanismo, transporte, navegación, tráfico, reglamento, señalización, manejo de siniestros, abasto combustible... [144].
- *Artes, cultura y entretenimiento*: Asistencia en la concepción de la actividad artística, cultural y diversión, como por ejemplo: música, preservación, videos... [145].
- *Instrumentos*: Diseño de métodos y creación de programas de cómputo para uso en CD que son objeto de propiedad intelectual ante IMPI e INDAUTOR [146].

## 4 Referencias

- [1] Stearns, P. N. (2018). *The industrial revolution in world history*. Routledge.
- [2] Ashton, T. S. (1997). *The industrial revolution 1760-1830*. OUP Catalogue.
- [3] Atkeson, A., & Kehoe, P. J. (2001). *The transition to a new economy after the second industrial revolution* (No. w8676). National Bureau of Economic Research.
- [4] Mowery, D. C. (2009). Plus ça change: Industrial R&D in the “third industrial revolution”. *Industrial and corporate change*, 18(1), 1-50.
- [5] Schwab, K. (2017). *The fourth industrial revolution*. Currency.
- [6] Beetham, H., & Sharpe, R. (Eds.). (2013). *Rethinking pedagogy for a digital age: Designing for 21st century learning*. Routledge.
- [7] Trilling, B., & Fadel, C. (2009). *21st Century Skills: Learning for Life in Our Times*. John Wiley & Sons.
- [8] Styazhkin, M. S. (2014). Managing competitiveness in XXI century: effective business instruments. *Journal of Economics and Social Sciences*, (5), 3.

- [9] McCarthy, John; Minsky, Marvin; Rochester, Nathan; Shannon, Claude (1955). A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence. Accessed on September 1, 2019: <http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html>
- [10] Norman, J. M. (Ed.). (2005). From Gutenberg to the Internet: A sourcebook on the history of information technology (Vol. 2). Norman Publishing.
- [11] Davis, M., Logemann, G., & Loveland, D. (1962). A machine program for theorem-proving. *Communications of the ACM*, 5(7), 394-397.
- [12] Taylor, W. (1967). Machine learning and recognition of faces. *Electronics Letters*, 3(9), 436-437.
- [13] Addis, T. R. (1956). Towards an "expert" diagnostic system. *ICL Technical Journal*, 1, 79-105.
- [14] Gentile, J. R. (1967). The first generation of computer-assisted instructional systems: an evaluative review. *AV Communication Review*, 15(1), 23-53.
- [15] Selfridge, O. G., & Neisser, U. (1960). Pattern recognition by machine. *Scientific American*, 203(2), 60-69.
- [16] Guzman-Arenas, A. (1968). Computer recognition of three-dimensional objects in a visual scene (No. MAC-TR-59). MASSACHUSETTS INST OF TECH CAMBRIDGE PROJECT MAC.
- [17] Oettinger, A. G. (1965). Automatic processing of natural and formal languages. *Information Processing*, 1.
- [18] McCulloch, W.S., Pitts, W.H. (1943) A logical calculus of the ideas immanent in nervous activity. *Bulletin of Mathematical Biophysics*, 5, 115-133.
- [19] Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6), 386.
- [20] Wiener, N. (1965). *Cybernetics or Control and Communication in the Animal and the Machine* (Vol. 25). MIT press.
- [21] Rechenberg, I. (1973). *Evolutionstrategie (Evolution Strategy)*. Frommann-Holzboog, Stuttgart.
- [22] Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). Intelligent decision making through a simulation of evolution. *Behavioral science*, 11(4), 253-272.
- [23] Holland, J.H. (1975) *Adaptation in Natural and Artificial Systems*. Ann Arbor, MI: University of Michigan Press.
- [24] Allen, J. P., & Van Buren, P. (1971). *Chomsky: selected readings*. London: Oxford University Press.
- [25] Newell, A. (1970). Remarks on the relationship between artificial intelligence and cognitive psychology. In *Theoretical approaches to non-numerical problem solving* (pp. 363-400). Springer, Berlin, Heidelberg.
- [26] Beni, G., Wang, J. *Swarm Intelligence in Cellular Robotic Systems*, Proceed. NATO Advanced Workshop on Robots and Biological Systems, Tuscany, Italy, June 26–30 (1989)
- [27] Zadeh, L. A. (1994). Fuzzy logic, neural networks, and soft computing. *Communications of the ACM*, 37(3), 77-85.
- [28] Kim, D., Goyal, A., Newell, A., Lee, S., Deng, J., & Kamat, V. R. (2019). Robotics, Automation, and Control. Semantic Relation Detection between Construction Entities to Support Safe Human-Robot Collaboration in Construction.

- [29] Rich, E. (1985). Artificial intelligence and the humanities. *Computers and the Humanities*, 19(2), 117-122.
- [30] Tukey, J. W. (1962). The future of data analysis. *The annals of mathematical statistics*, 33(1), 1-67.
- [31] Naur, P. (1974). *Concise survey of computer methods*. Petrocelli Books.
- [32] Press, G. (2013) A Very Short History of Data Science. *Forbes*, May 28.
- [33] Provost, F., & Fawcett, T. (2013). Data science and its relationship to big data and data-driven decision making. *Big data*, 1(1), 51-59.
- [34] Bu, Y., Howe, B., Balazinska, M., & Ernst, M. D. (2010). HaLoop: efficient iterative data processing on large clusters. *Proceedings of the VLDB Endowment*, 3(1-2), 285-296.
- [35] Stair, R., & Reynolds, G. (2013). *Principles of information systems*. Cengage Learning.
- [36] Elmasri, R., & Navathe, S. (2017). *Fundamentals of database systems*. Pearson.
- [37] Chen, Z., Xu, J., Tang, J., Kwiat, K., & Kamhoua, C. (2015, October). G-Storm: GPU-enabled high-throughput online data processing in Storm. In *2015 IEEE International Conference on Big Data (Big Data)* (pp. 307-312). IEEE.
- [38] Chen, W. Y., Song, Y., Bai, H., Lin, C. J., & Chang, E. Y. (2010). Parallel spectral clustering in distributed systems. *IEEE transactions on pattern analysis and machine intelligence*, 33(3), 568-586.
- [39] Dageville, B., Cruanes, T., Zukowski, M., Antonov, V., Avanes, A., Bock, J., ... & Lee, A. W. (2016, June). The snowflake elastic data warehouse. In *Proceedings of the 2016 International Conference on Management of Data* (pp. 215-226). ACM.
- [40] Laursen, G. H., & Thorlund, J. (2016). *Business analytics for managers: Taking business intelligence beyond reporting*. John Wiley & Sons.
- [41] Song, J., Guo, C., Wang, Z., Zhang, Y., Yu, G., & Pierson, J. M. (2015). HaoLap: a Hadoop based OLAP system for big data. *Journal of Systems and Software*, 102, 167-181.
- [42] Ho, T. B. (2016). -Knowledge Discovery. In *Knowledge Science* (pp. 70-93). CRC Press.
- [43] Oussous, A., Benjelloun, F. Z., Lahcen, A. A., & Belfkih, S. (2018). Big Data technologies: A survey. *Journal of King Saud University-Computer and Information Sciences*, 30(4), 431-448.
- [44] Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2017). Big Data and cloud computing: innovation opportunities and challenges. *International Journal of Digital Earth*, 10(1), 13-53.
- [45] Roiger, R. J. (2017). *Data mining: a tutorial-based primer*. Chapman and Hall/CRC.
- [46] Larson, D., & Chang, V. (2016). A review and future direction of agile, business intelligence, analytics and data science. *International Journal of Information Management*, 36(5), 700-710.
- [47] Yang, W. J. (2018). *Handbook of flow visualization*. Routledge.
- [48] Moor, J. (Ed.). (2003). *The Turing test: the elusive standard of artificial intelligence* (Vol. 30). Springer Science & Business Media.
- [49] Harnad, S. (1991) Other Bodies, Other Minds: A Machine Incarnation of an Old Philosophical Problem. *Minds and Machines*, 1(1):43–54, 1991.

- [50] Bringsjord, S., & Schimanski, B. (2003, August). What is artificial intelligence? Psychometric AI as an answer. In Proc. International Joint Conferences on Artificial Intelligence (pp. 887-893).
- [51] Morris, H. M. (1974). Scientific creationism. New Leaf Publishing Group.
- [52] Hortolà, P., & Carbonell, E. (2007). Creación versus evolución: del origen de las especies al diseño inteligente. *Asclepio*, 59(1), 261-274.
- [53] Simon, H. A., 1996. The Sciences of the Artificial. Third Edition ed. Cambridge, Massachusetts: MIT Press.
- [54] Brooks, R. A., & Maes, P. (Eds.) (1994). Artificial life. IV: Proceedings of the fourth international workshop on the synthesis and simulation of living systems (Vol. 4). MIT Press.
- [55] Gershenson, C., Trianni, V., Werfel, J., & Sayama, H. (2019). Self-Organization and Artificial Life. arXiv preprint arXiv:1903.07456.
- [56] Csanyi, V., & Csányi, V. (1989). Evolutionary systems and society: A general theory of life, mind, and culture. Duke University Press.
- [57] Maturana, H. R., & Varela, F. J. (1991). Autopoiesis and cognition: The realization of the living (Vol. 42). Springer Science & Business Media.
- [58] Von Bertalanffy, L. (1968). General system theory. *New York*, 41973(1968), 40.
- [59] Dell, P. F. (1982). Beyond homeostasis: Toward a concept of coherence. *Family Process*, 21(1), 21-41.
- [60] McLaughlin, P. (2000). What functions explain: Functional explanation and self-reproducing systems. Cambridge University Press.
- [61] Moore, E. F. (1962) Machine models of self-reproduction. In Proceedings of symposia in applied mathematics (Vol. 14, No. 1962, pp. 17-33). American Mathematical Society New York.
- [62] Neumann, J., & Burks, A. W. (1966). Theory of self-reproducing automata (Vol. 1102024). Urbana: University of Illinois Press.
- [63] Bell, M. A., & Deater-Deckard, K. (2007). Biological systems and the development of self-regulation: Integrating behavior, genetics, and psychophysiology. *Journal of Developmental & Behavioral Pediatrics*, 28(5), 409-420.
- [64] Jaynes, E. T. (1957). Information theory and statistical mechanics. *Physical review*, 106(4), 620.
- [65] Fetzer, J. H. (2012). Artificial intelligence: Its scope and limits (Vol. 4). Springer Science & Business Media.
- [66] Minsky, M. L. (1958, November). Some methods of artificial intelligence and heuristic programming. In Proc. Symposium on the Mechanization of Thought Processes, Teddington.
- [67] Granter, S. R., Beck, A. H., & Papke Jr, D. J. (2017). AlphaGo, deep learning, and the future of the human microscopist. *Archives of pathology & laboratory medicine*, 141(5), 619-621.
- [68] Warwick, K., & Shah, H. (2013). Good machine performance in Turing's imitation game. *IEEE Transactions on Computational Intelligence and AI in Games*, 6(3), 289-299.
- [69] Gams, M. (1954). Alan Turing-Einstein of computer science. *Informacijska Družba– IS 2012*, 1912, 7.
- [70] Besold, T., Hernández-Orallo, J., & Schmid, U. (2015). Can machine intelligence be measured in the same way as human intelligence?. *KI-Künstliche Intelligenz*, 29(3), 291-297.

- [71] Wang, P., & Goertzel, B. (Eds.). (2012). *Theoretical foundations of artificial general intelligence* (Vol. 4). Springer Science & Business Media.
- [72] Englemore, R. S., & Feigenbaum, E. (1993). Expert systems and artificial intelligence. *Expert Systems*, 100(2).
- [73] Dartnall, T. (Ed.). (2013). *Artificial intelligence and creativity: An interdisciplinary approach* (Vol. 17). Springer Science & Business Media.
- [74] Laird, J. E., Lebiere, C., & Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, 38(4), 13-26.
- [75] Liu, F., Shi, Y., & Li, P. (2017). Analysis of the Relation between Artificial Intelligence and the Internet from the Perspective of Brain Science. *Procedia computer science*, 122, 377-383.
- [76] Reid, M. (2016). Rethinking the Fourth Amendment in the Age of Supercomputers, Artificial Intelligence, and Robots. *W. Va. L. Rev.*, 119, 863.
- [77] Peltonen, E., Leppänen, T., & Lovén, L. (2019). EdgeAI: Edge-native Distributed Platform for Artificial Intelligence. 6G Wireless Summit, Levi, Finland.
- [78] Li, R., Zhao, Z., Zhou, X., Ding, G., Chen, Y., Wang, Z., & Zhang, H. (2017). Intelligent 5G: When cellular networks meet artificial intelligence. *IEEE Wireless communications*, 24(5), 175-183.
- [79] Serafini, L., & Bouquet, P. (2004). Comparing formal theories of context in AI. *Artificial intelligence*, 155(1-2), 41-67.
- [80] McCarthy, J. (1988). Mathematical logic in artificial intelligence. *Daedalus*, 297-311.
- [81] Miiller, Y. (1990). Decentralized artificial intelligence. *Decentralised AI*, 3-13.
- [82] Steels, L. (1993). The artificial life roots of artificial intelligence. *Artificial life*, 1(1\_2), 75-110.
- [83] Minsky, M. (2007). *The emotion machine: Commonsense thinking, artificial intelligence, and the future of the human mind*. Simon and Schuster.
- [84] Collins, A., & Smith, E. E. (Eds.). (2013). *Readings in cognitive science: A perspective from psychology and artificial intelligence*. Elsevier.
- [85] Chu, H. C., & Yang, S. W. (2012, March). Innovative semantic web services for next generation academic electronic library via web 3.0 via distributed artificial intelligence. In *Asian Conference on Intelligent Information and Database Systems* (pp. 118-124). Springer, Berlin, Heidelberg.
- [86] Kanal, L., & Kumar, V. (Eds.). (2012). *Search in artificial intelligence*. Springer Science & Business Media.
- [87] Davis, E., & Marcus, G. (2015). Commonsense reasoning and commonsense knowledge in artificial intelligence. *Commun. ACM*, 58(9), 92-103.
- [88] Arel, I., Rose, D. C., & Karnowski, T. P. (2010). Deep machine learning-a new frontier in artificial intelligence research. *IEEE computational intelligence magazine*, 5(4), 13-18.
- [89] Konar, A. (2018). *Artificial intelligence and soft computing: behavioral and cognitive modeling of the human brain*. CRC press.

- [90] Lemley, J., Bazrafkan, S., & Corcoran, P. (2017). Deep Learning for Consumer Devices and Services: Pushing the limits for machine learning, artificial intelligence, and computer vision. *IEEE Consumer Electronics Magazine*, 6(2), 48-56.
- [91] Bakhshipour, M., Ghadi, M. J., & Namdari, F. (2017). Swarm robotics search & rescue: A novel artificial intelligence-inspired optimization approach. *Applied Soft Computing*, 57, 708-726.
- [92] Helbing, D., Frey, B. S., Gigerenzer, G., Hafen, E., Hagner, M., Hofstetter, Y., ... & Zwitter, A. (2019). Will democracy survive big data and artificial intelligence?. In *Towards Digital Enlightenment* (pp. 73-98). Springer, Cham.
- [93] Gonzalez, R. T., Riascos, J. A., & Barone, D. A. (2017, November). How Artificial Intelligence is Supporting Neuroscience Research: A Discussion about Foundations, Methods and Applications. In *Latin American Workshop on Computational Neuroscience* (pp. 63-77). Springer, Cham.
- [94] Skilton, M., & Hovsepian, F. (2017). *The 4th Industrial Revolution: Responding to the Impact of Artificial Intelligence on Business*. Springer.
- [95] Özdemir, V., & Hekim, N. (2018). Birth of industry 5.0: Making sense of big data with artificial intelligence, "the internet of things" and next-generation technology policy. *Omics: a journal of integrative biology*, 22(1), 65-76.
- [96] Augusto, J. C. (2007). Ambient intelligence: the confluence of ubiquitous pervasive computing and artificial intelligence. In *Intelligent computing everywhere* (pp. 213 -234). Springer, London.
- [97] Demertzis, K., & Iliadis, L. (2015). A bio-inspired hybrid artificial intelligence framework for cyber security. In *Computation, Cryptography, and Network Security* (pp. 161-193). Springer, Cham.
- [98] Paradarami, T. K., Bastian, N. D., & Wightman, J. L. (2017). A hybrid recommender system using artificial neural networks. *Expert Systems with Applications*, 83, 300-313.
- [99] Hudson, V. M. (2019). *Artificial intelligence and international politics*. Routledge.
- [100] Bajaj, R., & Sharma, V. (2018). Smart Education with artificial intelligence based determination of learning styles. *Procedia computer science*, 132, 834-842.
- [101] Hamet, P., & Tremblay, J. (2017). Artificial intelligence in medicine. *Metabolism*, 69, S36-S40.
- [102] Bostrom, N. (2003). Ethical issues in advanced artificial intelligence. *Science Fiction and Philosophy: From Time Travel to Superintelligence*, 277-284.
- [103] Lin, P., Abney, K., & Jenkins, R. (Eds.). (2017). *Robot ethics 2.0: from autonomous cars to artificial intelligence*. Oxford University Press.
- [104] Nigon, J., Glize, E., Dupas, D., Crasnier, F., & Boes, J. (2016, July). Use cases of pervasive artificial intelligence for smart cities challenges. In *2016 Intl IEEE Conferences on UIC/ATC/ ScalCom/CBDCom/IoP/SmartWorld*. pp. 1021-1027. IEEE.
- [105] Agarwal, P. K., Gurjar, J., Agarwal, A. K., & Birla, R. (2015). Application of artificial intelligence for development of intelligent transport system in smart cities. *Journal of Traffic and Transportation Engineering*, 1(1), 20-30.
- [106] d'Inverno, M., & McCormack, J. (2015, June). Heroic versus collaborative AI for the arts. In *Twenty-Fourth International Joint Conference on Artificial Intelligence*.

- [107] Holland, W. (2003). U.S. Patent Application No. 10/001,847.
- [108] Chalmers, A. F. (2013). What is this thing called science? Hackett Publishing.
- [109] Harmanciogamalu, N. B., Singh, V. P., & Alpaslan, M. N. (Eds.). (2013). Environmental data management (Vol. 27). Springer Science & Business Media.
- [110] Gyamfi, A., & Williams, I. (Eds.) (2019). Big Data and Knowledge Sharing in Virtual Organizations. USA: IGI global. <https://doi.org/10.4018/978-1-5225-7519-1>.
- [111] Wersig, G., & Neveling, U. (1975). Terminology of documentation: A selection of 1200 basic terms. Paris: The UNESCO Press.
- [112] Harmanciogulu, N.B., Alpasian, N., & Singh, V.P. (1992) Design of water quality monitoring networks in R. N. Chowdhury (Ed), Geomechanics and Water Engineering in Environmental Management, pp. 267-296.
- [113] Harmanciogulu, N.B., & Alpasian, N. (1992) Water quality monitoring network design: a problem of multi-objective decision making, AWRA, Water Resources Bulletin 28(1) 179-192.
- [114] Kanehisa, M., Goto, S., Sato, Y., Kawashima, M., Furumichi, M., & Tanabe, M. (2013). Data, information, knowledge and principle: back to metabolism in KEGG. Nucleic acids research, 42(D1), D199-D205.
- [115] Poli, R. (2001). ALWIS. Ontology for knowledge engineers. Doctoral dissertation, University of Utrecht, the Netherlands.
- [116] Davis, G.B., & Olson, M.H. (1985). Management information systems. New York: McGraw Hill.
- [117] Kebede, G. (2010). Knowledge management: An information science perspective. International Journal of Information Management, 30(5), 416-424.
- [118] Zins, C. (2007). Conceptual approaches for defining data, information, and knowledge. Journal of the American society for information science and technology, 58(4), 479-493.
- [119] Kevin O Sullivan (2008) ICICKM 2008, 5th International Conference on Intellectual Capital, Knowledge. UK: Academic Published Limited Reading.
- [120] Provost, F., & Fawcett, T. (2013). Data Science for Business: What you need to know about data mining and data-analytic thinking. O'Reilly Media, Inc.
- [121] Van Der Aalst, W. (2016). Data science in action. In Process Mining (pp. 3-23). Springer, Berlin, Heidelberg.
- [122] Conway, D. (2003). Data science diagram. <http://www.drewconway.com/zia/2013/3/26/the-data-science-venn-diagram>
- [123] Schutt, R., & O'Neil, C. (2013). Doing data science: Straight talk from the frontline. O'Reilly Media, Inc.
- [124] Ozdemir, S. (2016). Principles of Data Science. Packt Publishing Ltd.
- [125] Hayashi, C. (1998). What is data science? Fundamental concepts and a heuristic example. In Data science, classification, and related methods (pp. 40-51). Springer, Tokyo.
- [126] Swan, M. (2013). The quantified self: Fundamental disruption in big data science and biological discovery. Big data, 1(2), 85-99.
- [127] Cao, L. (2017). Data science: challenges and directions. Communications of the ACM, 60(8), 59-68.

- [128] Lowrie, I. (2017). Algorithmic rationality: Epistemology and efficiency in the data sciences. *Big Data & Society*, 4(1),
- [129] Raschka, S. (2018). MLxtend: Providing machine learning and data science utilities and extensions to Python's scientific computing stack. *J. Open Source Software*, 3(24), 638.
- [130] Baesens, B. (2014). *Analytics in a big data world: The essential guide to data science and its applications*. John Wiley & Sons.
- [131] Cielen, D., Meysman, A., & Ali, M. (2016). *Introducing data science: big data, machine learning, and more, using Python tools*. Manning Publications Co.
- [132] Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., ... & Kumar, V. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, 29(10), 2318-2331.
- [133] Power, D. J. (2016). Data science: supporting decision-making. *Journal of Decision systems*, 25(4), 345-356.
- [134] Waller, M. A., & Fawcett, S. E. (2013). Data science, predictive analytics, and big data: a revolution that will transform supply chain design and management. *Journal of Business Logistics*, 34(2), 77-84.
- [135] Baesens, B., Van Vlasselaer, V., & Verbeke, W. (2015). *Fraud analytics using descriptive, predictive, and social network techniques: a guide to data science for fraud detection*. John Wiley & Sons.
- [136] Taylor, L., Floridi, L., & Van der Sloot, B. (Eds.). (2016). *Group privacy: New challenges of data technologies* (Vol. 126). Springer.
- [137] Biem, A., Butrico, M., Feblowitz, M., Klinger, T., Malitsky, Y., Ng, K., ... & Sow, D. (2015, March). Towards cognitive automation of data science. In *Twenty-Ninth AAAI Conference on Artificial Intelligence*.
- [138] Chen, Y. C., & Hsieh, T. C. (2014). Big data for digital government: Opportunities, challenges, and strategies. *International journal of public administration in the digital age (IJPADA)*, 1(1), 1-14.
- [139] Klačnja-Milićević, A., Ivanović, M., & Budimac, Z. (2017). Data science in education: Big data and learning analytics. *Computer Applications in Engineering Education*, 25(6), 1066-1078.
- [140] Adam, N. R., Wieder, R., & Ghosh, D. (2017). Data science, learning, and applications to biomedical and health sciences. *Annals of the New York Academy of Sciences*, 1387(1), 5-11.
- [141] Fairfield, J., & Shtein, H. (2014). Big data, big problems: Emerging issues in the ethics of data science and journalism. *Journal of Mass Media Ethics*, 29(1), 38-51.
- [142] Hofmann, M., Neukart, F., & Bäck, T. (2017). Artificial intelligence and data science in the automotive industry. *arXiv preprint arXiv:1709.01989*.
- [143] Wu, J., Guo, S., Li, J., & Zeng, D. (2016). Big data meet green challenges: Big data toward green applications. *IEEE Systems Journal*, 10(3), 888-900.
- [144] Dobre, C., & Xhafa, F. (2014). Intelligent services for big data science. *Future Generation Computer Systems*, 37, 267-281.
- [145] Manovich, L. (2015). Data science and digital art history. *International Journal for Digital Art History*, (1).

[146] Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., & Witten, I. H. (2009). The WEKA data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1), 10-18.