Neural Computing and Applications (2025) 37:7575–7607
https://doi.org/10.1007/s00521-023-08328-z

S.I.: KNOWLEDGE SERVICES: CHALLENGES AND FUTURE PROSPECTS

A systematic review of integrated information theory: a perspective from artificial intelligence and the cognitive sciences

Luz Enith Guerrero[1] · Luis Fernando Castillo[1] · Jeferson Arango-López[1] · Fernando Moreira[2]

Received: 13 September 2022 / Accepted: 16 January 2023 / Published online: 7 February 2023

Abstract

The study of consciousness has gained momentum in recent years by the scientific community. In this same sense, the relationship between cognitive sciences and artificial intelligence presents a fundamental theoretical framework in the study of integrated information theory (IIT) as a theory that makes its way into the knowledge of consciousness. However, there are few studies that integrate these topics and a systematic review of the literature is highly pertinent. This paper seeks to identify methods, methodologies or computational solutions using artificial intelligence and cognitive science fundamentals that can provide some kind of solution to the challenges posed by IIT.

Keywords IIT · Integrated information · Artificial intelligence · Consciousness · Systematic review · Cognitive systems

1 Introduction

In the history of mankind, there has been a close relationship between mind and body. However, there has been no common understanding to define consciousness. For this reason, there are many definitions of consciousness. For example, it is presented as "the ability of human beings to assign special feelings to a mental capacity" or "subjecting the mental state to past, present and future scenarios with reflection on themselves as aware of the surrounding environment" [1].

At the end of the last century, neuroscientists such as Crick [2] and physicists such as Penrose [3] claimed that it was time for science to address consciousness and a veritable explosion of scientific work on the subject was generated. According to Chalmers [4], a science of consciousness is an impossibility for some, because by nature science is objective and consciousness is subjective. In this sense, a question that arises is whether a science of consciousness could exist, and in order to give an answer, several theories have been presented that seek to define consciousness. In this study, in particular, consciousness is approached within the framework of the IIT This theory was put forward by the neuroscientist Giulio Tononi [5] as an effort to make sense not only of the philosophy of consciousness, but also of conscious experiences and the neurology of consciousness. Moreover, it is done in a rigorous way within a theoretical framework that defines consciousness, quantifies it, and makes it possible to determine which systems are conscious and which are not.

Neuroscience has made progress in understanding how brain mechanisms perform cognitive functions as diverse as decision-making, language analysis, motor control, etc. An important problem in neuroscience is to find which groups of components of a system form irreducible sets with shared functionality. This problem is approached from the perspective of IIT, which starts from two fundamental postulates about subjective experience: that it contains information and that this information is irreducibly integrated. The integration of information is the key principle for understanding IIT, in which it is understood that the

✉ Fernando Moreira
fmoreira@mail.upt.pt

Luz Enith Guerrero
luzenith_g@ucaldas.edu.co

Luis Fernando Castillo
luis.castillo@ucaldas.edu.co

Jeferson Arango-López
jeferson.arango@ucaldas.edu.co

1 Departamento de Sistemas e Informática, Universidad de Caldas, Manizales, Colombia

2 REMIT, IJP, Universidade Portucalense and IEETA, Universidade de Aveiro, Aveiro, Portugal

more interconnected or integrated a system is, the more conscious it is. Evaluating the irreducibility of a system consists of finding the minimum information partition (MIP). This is a numerically intractable problem due to the combinatorial explosion generated by the number of partitions that result [6].

Consequently, given the importance that, not only for neuroscience entails this problem, but for other areas of knowledge, we intend to review the possible relationships between cognitive sciences, artificial intelligence and IIT that propose solutions to the challenges faced by the theory. All this from the systematic review of the literature and subsequent analysis of the most important findings.

The rest of this paper is distributed as follows: Sect. 2 presents a contextualization of the research topics. Section 3 describes the protocol and research questions of this study. Section 4 shows the analysis of the results of the extracted data and the answers to the research questions. A discussion of future work and challenges is found in Sect. 5, and finally, Sect. 6 ends with a summary of the conclusions.

2 Background

In this section, descriptions of artificial intelligence, cognitive systems and integrated information theory will be briefly presented as a central basis for identifying possible relationships between these topics.

2.1 Artificial intelligence (AI)

According to Nilsson [7], one of the pioneers in the field, "artificial intelligence is that activity devoted to making intelligent machines, and intelligence is that quality that enables an entity to function properly and with foresight in its environment". Holland and Gamez [8] note that "A system with general artificial intelligence is capable of performing a wide range of human-like intelligence that is not limited to a single task, and systems with artificial intelligence are capable of intelligent reasoning and behavior and often have some form of learning mechanism".

Some ideas inspired by ancient Greek philosophy claim that only humans are capable of conceptual thinking, as opposed to perceptual thinking, which is in higher animals, some machines and of course humans. It is not yet known whether AI systems can engage in conceptual thinking [9].

For [10], artificial general intelligence (AGI) has as its main goal to enable a machine to do general intelligence tasks as humans do. This could be achieved in theory through approaches such as machine learning and cross-domain optimization. The problem is that this current AI focuses on specific domains employing machine learning. So, the concept of consciousness is very interesting to achieve both approaches because it not only encodes, but at the same time processes various information with seamless integration. If you want a machine to develop its own rules it will require the concept of a learning machine and in turn, it has been said that, for this, a machine needs consciousness. In addition, because of the development in this field, more efficient computational applications are required in information processing so that they can learn from the environment, and this is in turn an impetus for the development of AI mainly through machine learning and neural networks [11].

The foundations of AI rest on several disciplines spanning philosophy, logic, biology, statistics, engineering and to which neuroscience is added.

2.1.1 Computational neuroscience

Computational neuroscience is defined as the area in which problems lie simultaneously within computational science and neuroscience [12]. Its main objective is to virtually represent brain neural networks and their interactions in order to understand how the functions that allow us to perceive, process and react to stimuli, among others, come from electrochemical communication between individual neurons. The idea is to generate computer simulation models of neuronal activity using mathematical models based on statistical estimates and thus be able to appreciate brain activity in other ways. The data obtained from experiments with neuroimaging and other techniques are used for models and simulations to make predictions about the functions and networks involved [13]. Computational neuroscience also makes it possible to establish a similarity between the way living organisms learn and computerized forms of learning or machine learning [14].

Current neuro computational efforts commonly posit that the goal is to understand consciousness and cognition in sufficiently complex biological networks [15].

According to [11], cognitive sciences have developed computational models that bring cognition into functional components. For its part computational neuroscience has modeled how the interaction of neurons can implement components of cognition. Thus, the intersection of cognitive science, computational neuroscience and AI is moving toward brain-based computational models. Such models allow mimicking brain information processing during perceptual, cognitive and control tasks and can be tested with brain and behavioral data. In particular, this interaction between neurons has been modeled as artificial neural networks, which are one of the many types of structures used for computational learning.

2.1.2 Artificial neural networks

An artificial neural network (ANN) is a collection of interconnected units, and its properties are determined by its topology and the properties of the "neurons" or units used in such networks [16]. AI problems with approaches involving ANNs and statistical techniques were termed "non-symbolic", with application mainly in pattern recognition, speech processing and computer vision [7].

Brain neural networks have heterogeneous nodes coupled with forward and backward feedback. This feedback occurs with a variety of spiral neurons that present connections with synapses of different classes and with short- and long-term dynamics that constrain computational models. Experiments have shown that brain networks are constantly changing. Many neuroscience models take probabilistic calculations, and the stochastic form of spiking neurons allows the network to solve problems heuristically. Thus, a state of the network is, for example, the representation of a possible solution to the problem. This suggests a possible relationship between flickering internal states of brain networks and response in terms of perception and behavior [15].

There has been recent work on two classes of networks: deep convolutional and recurrent. The former have enabled advances in image recognition while recurrent networks have done so in natural language processing, and have performed well with time series problems such as machine translation [17]. Currently, there have been many publications on machine learning to train ANNs that implement algorithms for tasks such as data classification [15].

2.1.3 Machine learning

Machine learning (ML) methods are playing a very important role in AI, in tasks such as classification, regression in application areas such as computer vision, speech recognition, bioinformatics, and big data analytics [7]. Machine learning leverages both mathematics and statistics with the idea of automating tasks that are complex in various fields [18]. The idea is that the machine can adapt to new circumstances, detect and extrapolate patterns.

There are several kinds of machine learning: Unsupervised learning, whose most common task is clustering, which tries to detect potentially useful clusters from the input examples. This technique seeks to discover similarities between the data and from them perform some classification [7]. Another kind of learning is supervised learning, where a function is learned from an example as an input–output pair, that is, the machine learns a function that relates input and output. And finally, semi-supervised learning which takes a few labeled examples and must do what it can from a large number of unlabeled examples. Deep learning can be seen mainly in multiple tasks, such as object recognition in images, transcription of speech into text, searches according to user interests in publications, news etc., [19]. In deep learning we have representation-learning methods with multiple levels of representation. Which are achieved by composing simple, nonlinear modules that take the representation that starts with raw data into a representation at a higher level. In this learning, machines learn from both experience and knowledge without the need for explicit programming and with raw data to extract useful patterns.

New statistical machine learning methods such as "multivariate pattern analysis" generate data-driven stimulus response models employing MRI that impact human cognition [20].

2.2 Cognitive systems

Cognition is "the mental action or process of acquiring knowledge and understanding thought, experience, and the senses," as well as "a perception, sensation, idea, or intuition resulting from the process of cognition" [15]. The development of computational modeling led to the creation of cognitive science, and this has provided clues for AI researchers. For its part, AI has helped cognitive science with recent concepts to understand the workings of the mind [7]. A shift in thinking has been generated in recent times in cognitive science due to the Bayesian approach. This new approach suggests that the human (or animal) mind is a probabilistic machine in a certain way and is not a faulty machine. According to this view of human cognition, "we constantly estimate and revise probabilities of events in the world, taking into account any new information, and more or less following probabilistic (including Bayesian) rules" [21]. Within the cognitive sciences, the idea of treating minds as virtual reality machines (artificial cognitive systems) is gaining momentum.

Another aspect is metacognition or "some cognitive process or structure about another cognitive process or structure" [22]. These are higher-order thinking skills that include knowledge about when and how to use strategies to learn or solve problems depending on particulars [23]. Metacognition has been addressed by computational frameworks for cognition with different approaches according to neuropsychological and psychological theories used in its development [22].[1] On the other hand, if consciousness and cognition are concerned, for [24] the separation of both has been fundamental for the scientific

[1] These frameworks or cognitive architectures refer to the theory of mind structure, taking human consciousness, with the aim of bringing cognitive knowledge into computational models [1].

study of consciousness. While it is true, AI systems can be somewhat compared to human brain functions, none of the systems show advanced cognitive functions that resemble human consciousness and emotions. However, one theory of consciousness, IIT, proposes that consciousness can exist wherever information can be reasonably processed, whether brain or machine [17].

2.3 Integrated information theory of consciousness

Reference [25] defines consciousness as the ability to experience something more than one's internal physical underpinnings, as happens in humans. It is an empirical fact that when we are conscious, we experience, for example, the external world rather than what happens inside our brain. [24] proposes two kinds of consciousness, phenomenal or qualia (raw experiences) and access consciousness (information in human minds that is accessed to control behavior, reasoning, decision making) [1]. The focus of consciousness science in recent years was the search for correlations between brain areas and states of consciousness. Questions about what certain brain areas do and what they correlate with according to [4] are the "easy problems" of consciousness and do not address the "hard problem of consciousness" posed by the question Why is every physical process in the brain accompanied by consciousness?

IIT is an evolving theoretical framework that seeks to explain what a system requires for consciousness to arise, proposed by [5] has had several revisions, one in 2008 [26] and the most recent in 2014 [6]. IIT posits the idea that not only humans are the only ones capable of sensing and self-awareness.[2] IIT is derived from mental experiments leading to phenomenological axioms and ontological postulates [5]. Thus, a system comprises a set of elements that can assume discrete states, and according to the interaction between the elements there will be a specification of the rules of transition between these states.

IIT is based on two fundamental postulates about subjective experience: it is differentiated and irreducibly integrated [27]. IIT relates the amount of integrated information to the degree of consciousness [28] and establishes a way to measure the consciousness of a system. That mathematical measure is known as phi, Φ. Calculating Phi is a demanding task not only experimentally but mathematically and computationally because of the enormous amount of computations and is only functional in very small systems [29]. This is probably one of the main theories currently in consciousness science that relies on a mathematical framework to assess the quality and quantity of consciousness.

[2] Reference [28] believe that part of the difficulty in dealing with the difficult problem of consciousness has been the way science has approached it, so the theory aims to start from the simple or primary consciousness and asks what kind of physical mechanisms could give rise to it.

3 Research method

A systematic literature review is a method that allows studying, evaluating and interpreting studies relevant to a particular research question, a phenomenon of interest or specific area [30]. This systematic review was conducted taking 2008 as a reference point because it was the year of publication of the second version of the ITI. For this research, a rigorous protocol was followed taking as guidelines what was proposed by [31, 32] for the development of mapping studies and by Kitchenham and Charters [33] for the development of systematic literature reviews. The publications obtained are used to answer the research questions posed. As a result of this systematic review, an overview of the research is presented taking into account several aspects: (a) analysis of the information according to demographic characteristics such as type of publications, predominant researchers, publications by year; (b) identification of different contributions from AI and cognitive sciences to IIT. The mapping process was divided into several stages as shown in Fig. 1.

3.1 Identification of the need for a review

According to [33], existing systematic reviews of the topic of interest should be identified and studied according to the appropriate evaluation criteria. For this and as suggested by [34], first, in order to broaden the scope of the results, other systematic literature studies in the area of integrated information theory were sought, and a procedure similar to the one followed in the main search of this systematic review was carried out, using the same digital libraries presented in Table 1.

To perform appropriate searches on the related terms, the following search string was designed:

("Integrated information theory" OR IIT OR "Integration information theory") AND ("review" OR "systematic mapping" OR "state of the art").

As a result of this search, no secondary studies were identified, so it is considered pertinent to conduct a systematic mapping of IIT and its relationship with artificial intelligence and cognitive systems as a need to find references in the literature to provide a basis for those interested in delving deeper into these topics.

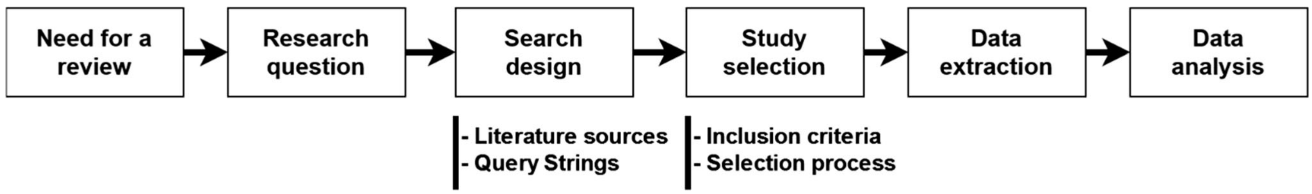


Fig. 1 Stages of a systematic review

Table 1 Digital libraries

Library	Link
ACM digital library	www.dl.acm.org
IEEE Xplore digital library	www.ieeexplore.ieee.org
Sciencedirect	www.sciencedirect.com
Scopus	www.scopus.com
SpringerLink	www.springerlink.com
Nature	www.nature.com.ezproxy.unal.edu.co/
Engineering-database	https://search-proquest-com.ezproxy.unal.edu.co/engineeringjournals/
Web of science	www.webofscience.com
Willey online library	www.onlinelibrary.wiley.com

3.2 Research questions

The objective of this study is to find and analyze the state of the art in integrated information theory and its relationship with artificial intelligence and cognitive systems. The objective has been divided into two research questions that will allow for a more productive review. Through the questions, it is intended to identify the current state of IIT, as well as to make an exploration in order to find and understand the underlying relationships with artificial intelligence and cognitive sciences, which present solutions or alternatives to the challenges posed by IIT. The following research questions have been formulated (see Table 2).

3.3 Search design process

Systematic mappings use automatic searches performed on scientific databases and manual navigation by reviewing specific journal and conference papers in the research field. The objective of this search process was to find, as far as possible, the main studies related to the search questions using non-biased search strategies.

3.3.1 Literature sources

To ensure that the relevant sites were considered, bibliographic references were selected from the databases mentioned in Table 1. In that order of ideas, mainly databases oriented to engineering, computational sciences and also general databases such as Nature and Springer were chosen to cover relevant material from studies on integrated information theory, cognitive sciences and artificial intelligence. In addition, the following were considered:

Journal of Consciousness.

Journal of Artificial Intelligence and Consciousness (2020 Vol 1 Issue 1 onward). First known as: International Journal of Machine Consciousness (2009 Vol. 1 Issue 1—2014 Vol. 6 Issue 2).

International Workshop on Machine Consciousness.

Conference: IEEE symposium on computational intelligence and games.

Conference: International Work-Conference on the Interplay Between Natural and Artificial Computation.

Conference: IEEE International Conference on Cognitive Informatics.

International Conference on Cognitive Systems, CogSys.

Conference on Brain-Inspired Cognitive Systems BICS.

3.3.2 Query strings

In order to obtain all the results related to the research questions, the query string is formulated considering [33]. Since they proposed to specify population, intervention, comparison and outcome. In this case, the focus will be on the population dimension and since the interest is in three areas simultaneously, the search chain was posed as a conjunction of the corresponding populations, thus:

Population: "integrated information theory" AND population "Artificial intelligence" AND Population "Cognitive Sciences".

Table 2 Research questions

Research question	Interest and motivation	Subquestions
RQ1. ¿What are the characteristics of the studies that relate to the main objective of this systematic review?	Identifying the type of publication, such as journal publications, is important because it is an indicator of maturity in an emerging field of research [35]. In addition, the behavior of the number of publications in different years is a reference of how the activity of a research field changes [36]. Information on the geographic distribution of publications is relevant to understand how the concept of IIT and its relationships according to this study, extends geographical boundaries Finally, the predominant researchers in IIT are important to identify the key authors in the area	RQ1.1 ¿What types of sources are the articles mostly published in? RQ1.2 ¿How has the number of publications evolved over the years? RQ1.3 ¿How are the publications geographically distributed? RQ1.4 ¿Who are the predominant researchers? RQ1.5 ¿How are publications distributed between academia and industry? RQ1.6 ¿What types of publications are published?
RQ2. ¿What are the relationships between IIT, artificial intelligence and cognitive systems?	To identify which techniques, methods, etc., from AI and cognitive systems are proposed in the literature to address the problem of consciousness and the challenges in this regard, within the framework of IIT	RQ2.1 ¿What types of relationships, if any, are present between IIT, artificial intelligence, and cognitive systems? RQ2.2 ¿What applications exist that use or take IIT as a reference? RQ2.3 ¿ What solutions are presented concerning the challenges posed for the development and/or application of IIT? RQ2.4 ¿What trends are there in relation to the development of ITI and what future work is planned? RQ2.5 ¿What barriers are encountered to the advancement of ITI? RQ2.6 ¿What criticisms or problems are raised about IIT?

Due to the existence of several terms used to name the defined populations, the resulting string was:

("Integrated information theory" OR IIT OR "Integration information theory") AND ("Artificial intelligence" OR "AI" OR "Deep learning" OR "Machine learning" OR "Neural networks" OR "Case based reasoning" OR "Expert systems" OR "agents") AND ("Cognitive science" OR cognition).

When running this query string and considering only publications since 2008, very few publications were obtained. For this reason, the third area was expanded to include cognitive sciences or general information on consciousness.

3.4 Studies selection

This studies selection strategy was done based on the steps proposed according to [31, 33, 34]. Articles based on titles and abstracts are excluded, as well as full-text reading.

3.4.1 Inclusion and exclusion criteria

The following criteria were used for the selection of relevant publications:

- English language publications only.
- Publications published between 2008 and 2021

We excluded panels, conference prefaces and special issues, book reviews, news, short articles (less than 4 pages) and PhD symposium papers (i.e., publications without bibliographic information, papers that only report work in progress, and non-peer-reviewed publications).

3.4.2 Stage 1—automatic search

The digital libraries used in this study are presented in Table 1, which cover the publication sources described in Sect. 3.3.1. For this purpose, a search was performed on each of the digital libraries and the references found were stored in bibliographic files. Initially, 1012 resources were identified. The distribution according to data source is shown in Fig. 2.

Libraries	Number of publications
ACM	2
Engineering DB	67
Springer	267
Nature	31
IEEExplorer	136
Science direct	102
Scopus	78
Web of science	24
Wiley	217
Journal Consciousness	83
Otros	5
Total	1012

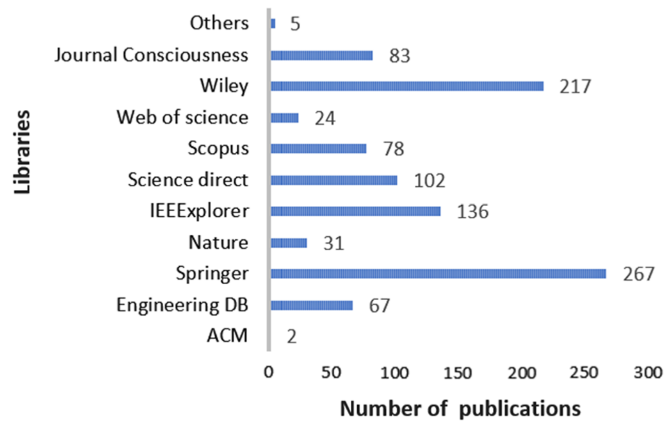


Fig. 2 Results for each library

3.4.3 Stage 2—removal of duplicates

All references were imported into a reference management system, and duplicate references were automatically removed. After this, the list of articles was manually reviewed for duplicate references that had not been detected by the handler and 97 duplicate publications were excluded. Additionally, 32 publications from years prior to 2008 were detected and eliminated, leaving a total of 883 references.

3.4.4 Stage 3—titles and abstracts

To identify publications related to IIT, artificial intelligence and cognitive systems, all titles and abstracts were reviewed according to the inclusion and exclusion criteria defined for each of the references. After this stage, 785 references were discarded leaving 98 publications.

3.4.5 Stage 4—quick reading

Each publication from the previous stage was reviewed for both results and conclusions and briefly reviewed for content and all papers that did not reflect the themes of the study, did not address any of the research questions or were delta papers were excluded. Finally, 58 publications were selected.

3.4.6 Stage 5—secondary studies

No secondary studies were identified for the topic addressed.

3.4.7 Stage 6—manual search

The search of the digital libraries was complemented with some manual searches in order to ensure the widest possible coverage. Four publications were identified using this type of search.

3.4.8 Stage 7—the book The Blackwell companion consciousness is related

For this book, a review of the chapters was done applying stages 2, 3 and 4.

3.4.9 Final result

Finally, this review included 63 relevant publications. A summary of the number of publications resulting in each stage according to the study review process is presented in Fig. 3. For a complete list of the publications and their codes, please refer to "Appendix 1".

3.5 Data extraction

Using the Atlas.ti® tool, the qualitative analysis of the data was performed considering what was suggested by [34, 37] to obtain information from the studies reviewed. The stages of this qualitative analysis were as follows:

- Data processing and preparation: the Atlas.ti® tool was used to group the final publications resulting from systematic mapping.
- Codes and labels with a symbolic meaning are established in order to categorize the data so that they can be grouped with respect to the specific research questions posed [37].

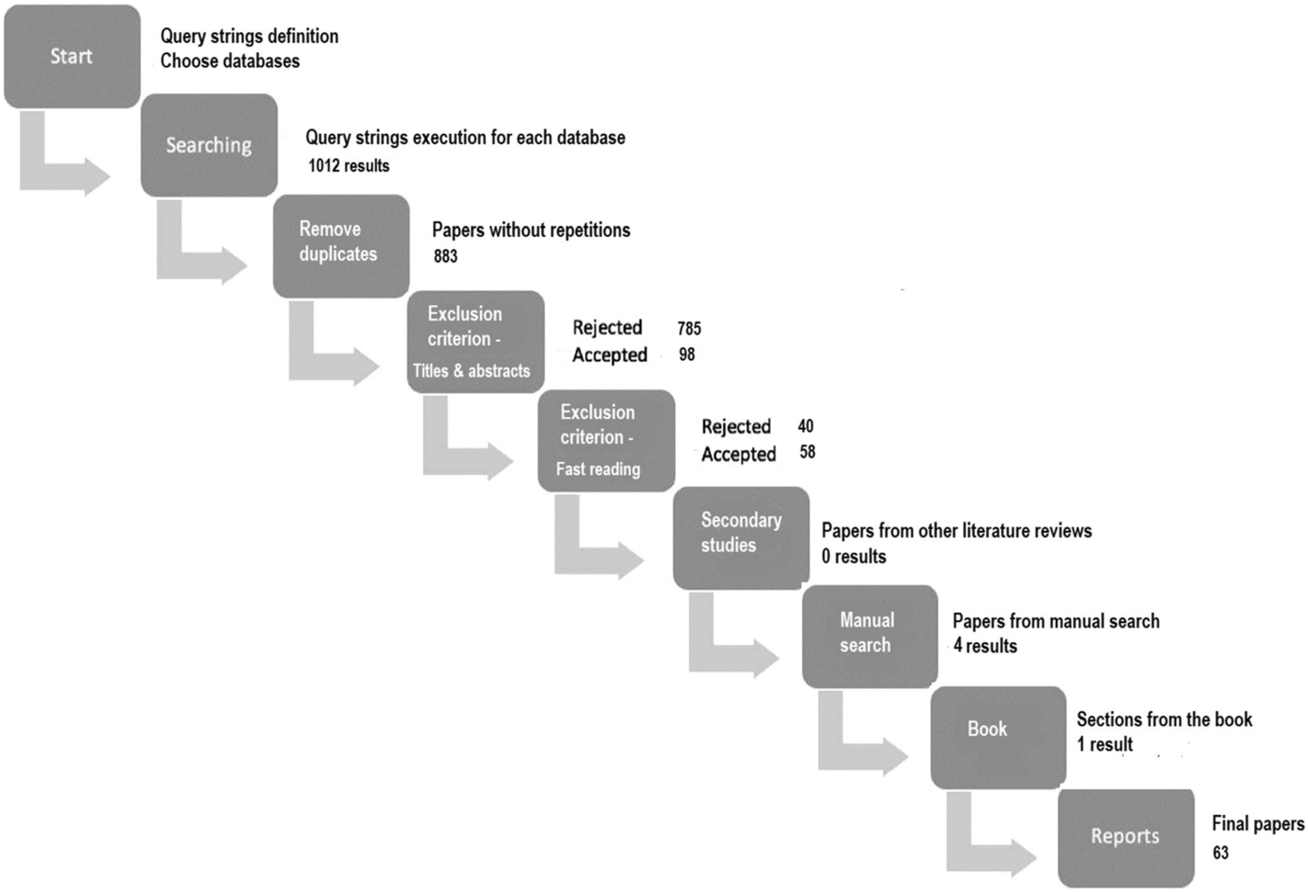


Fig. 3 Stages of the initial review process

- Code groups are established. The list of codes is sought to be grouped into a smaller number of categories, so that they function as explanatory or inferential codes used to explain, identify an emerging theme or configuration [37]. These categories will be presented later in the results section.
- Data visualization: Visual representation of the data is essential for a better understanding of the data, which is why the results will be presented through the use of tables and graphs.

In order to extract data from the classified primary studies, a spreadsheet template was developed with the following categories: author and affiliation (organization and country), title, year of publication, type of publication (journal article, conference proceedings and book chapter), type of research (empirical research (R), experience report (E) and both opinion and theoretical papers (N)), the type of contribution of the article: method, model, methodology, framework, algorithm, approach, technique, application or other, the degree of relationship with IIT (total, partial, none).[3] The classifications and categories considered by [38, 39].

3.6 Data analysis

The information extracted for each item was tabulated and is presented in Sect. 4. The data that were tabulated to answer the research questions are presented below:

Number of papers by source; number of publications by year; number of publications by country; number of publications by type of research; the number of publications presenting solutions to IIT challenges, the field of application of IIT (is the field in which IIT is used in some way).

4 Results

This section summarizes the results obtained from the data mining process.

RQ1. What are the demographic characteristics of the studies?

[3] Total refers to whether the IIT is used. Partial refers to whether it only uses some features or components of ITI. None refers to not using any of the theory.

To answer this question, several sub-questions were posed that will be answered below and for which the process defined in Sect. 3.5 was applied.

RQ1.1 In what type of sources are the articles mainly published?

Figure 4 shows the distribution of the 63 publications, with journal articles (24) being the most relevant publications.

RQ1 2 How has the number of publications evolved over time?

Primary studies were reviewed between 2008 and 2021. Figure 5 shows the number of studies published per year and Fig. 6 likewise, but discriminating according to the type of relationship.

RQ1.3 How are the publications distributed geographically?

To determine the geographical distribution of the publications, the country of affiliation of the first author was taken, with Europe being the continent with the highest participation and then North America. Figure 7 shows the geographical distribution according to the above considerations.

RQ1.4 Who are the predominant researchers?

Figure 8 illustrates the authors who publish the most of papers; specifically, the figure shows authors with more than one publication. The three main authors and those authors who are not among the three main authors but who are involved in more than one publication are considered in this relationship. “Appendix 3” presents the complete list of authors.

4.1 ¿RQ2 Which relationships are present between IIT, artificial intelligence and cognitive systems?

Three dimensions of those described in Sect. 3.6 have been considered to present some of the results of this systematic review. These dimensions are: type of contribution, type of research, and degree of relationship with IIT, which are shown in Fig. 9, according to the guide for bubble diagram charts in [32]. In the systematic review for the classification of papers according to the type of relationship, only the one that was most relevant according to the research was considered, this given that several schemes could be developed in one study, such as presenting a method and a framework. Thus, each publication studied was associated with a single relationship.

Finally, in order to answer this question, several sub-questions were posed that will be answered below and for which the process defined in Sect. 3.5 was applied.

RQ2.1 What types of relationships, if any, are present between IIT, artificial intelligence and cognitive systems?

Specifically, for those publications that presented explicit relationships between cognitive systems, AI and IIT, the types of relationship or contribution that were considered were: method, model, methodology, algorithm, applicative, technique and framework, and for those publications that did not show direct relationships with IIT, the category, other, was considered, and this category was subdivided into four subcategories: OTC, if the publication made reference or another theory of consciousness; Criticism, if the publication presented a critical analysis of IIT, and theoretical, if the publication was a study from the theoretical foundation (see Fig. 10).

The different Tables 3, 4, 5, 6, 7, 8, 9 presented below are evaluated according to these types of relationship and also show whether these studies are critical toward IIT and/or whether they propose some kind of solution for the materialization of the theory:

Thirty-nine publications were found that presented some kind of relationship between IIT, AI and cognitive systems, as shown in Fig. 10.

IIT states that a theory of consciousness has to consider the essential properties of one’s own experience of phenomena in the face of which [39] presents a theoretical framework that considers such experiences as self-representations of knowledge, and which [22] show as a connection between theories of consciousness and analytical techniques from cognitive science and AI, in their study on artificial systems implementing metacognitive components.

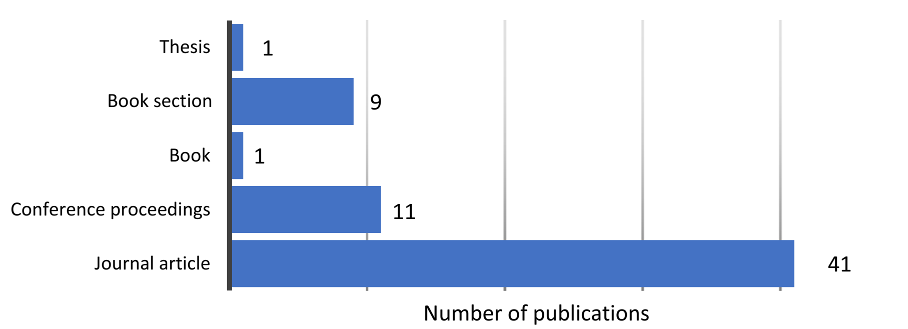
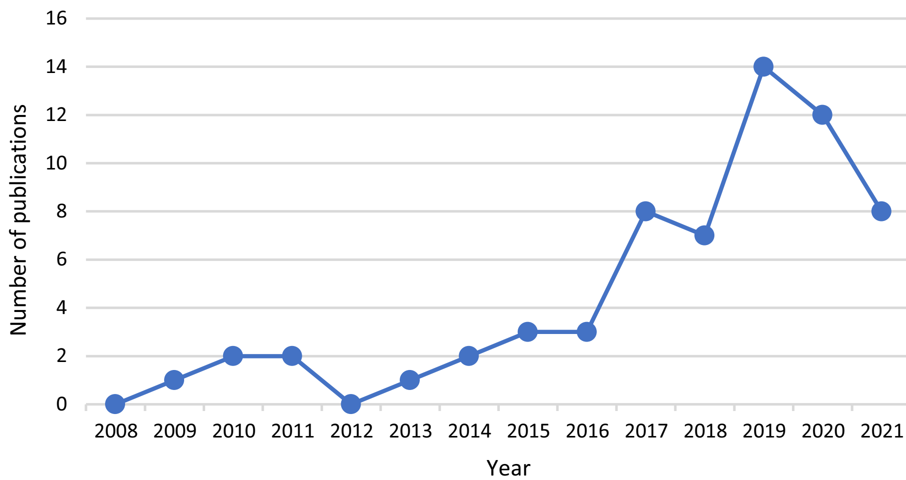


Fig. 4 Type of publications

Fig. 5 Publications per year



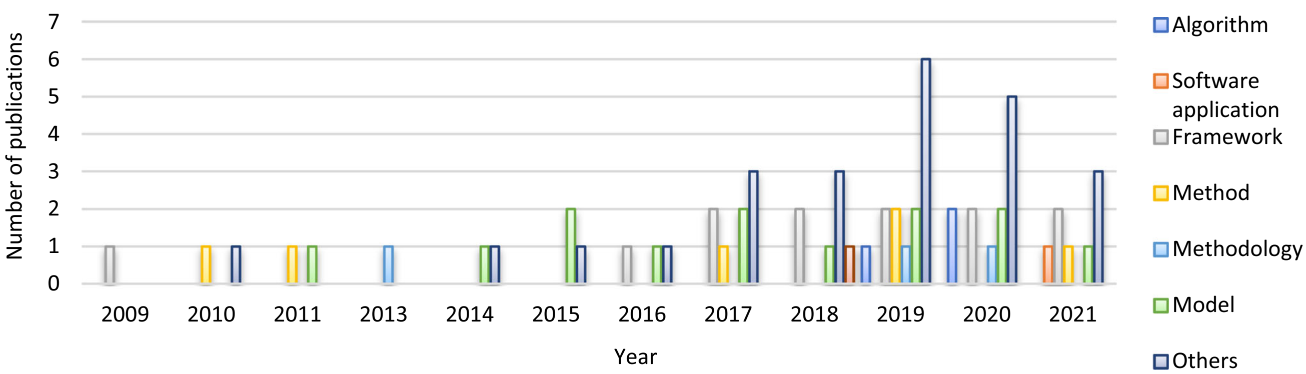


Fig. 6 Publications per year by type of relationship

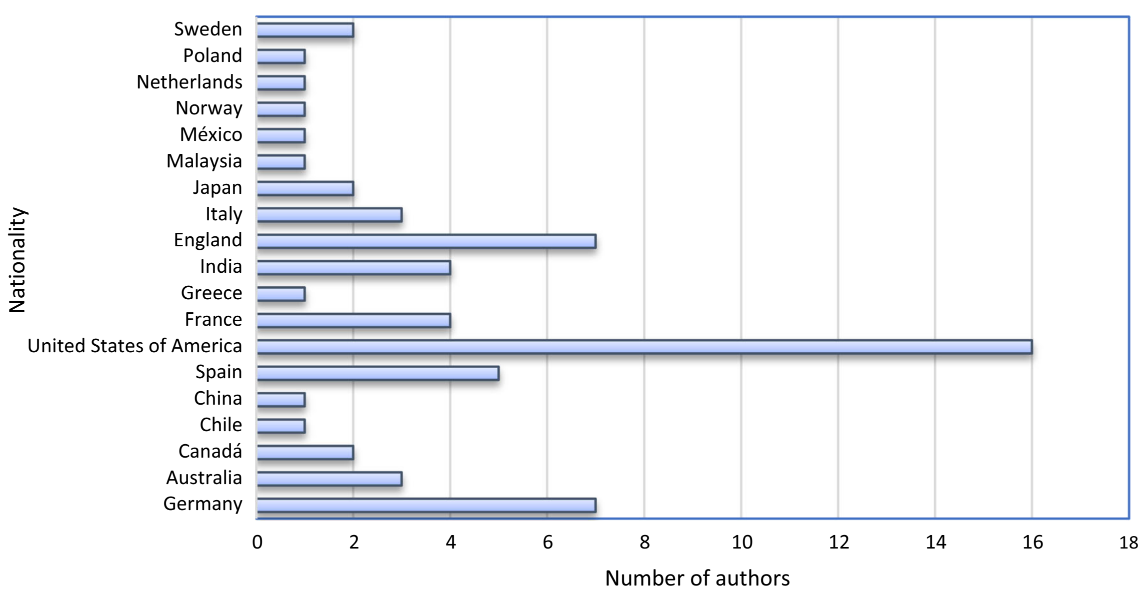


Fig. 7 Geographic distribution

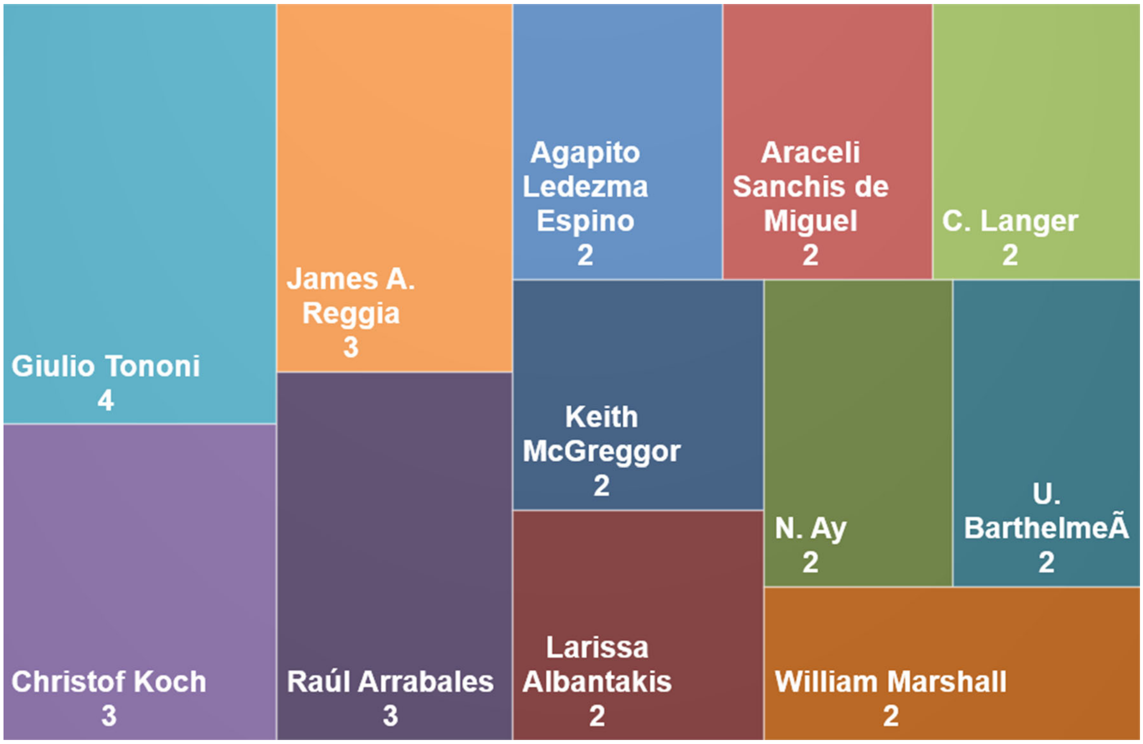


Fig. 8 Predominant researchers with more than one publication

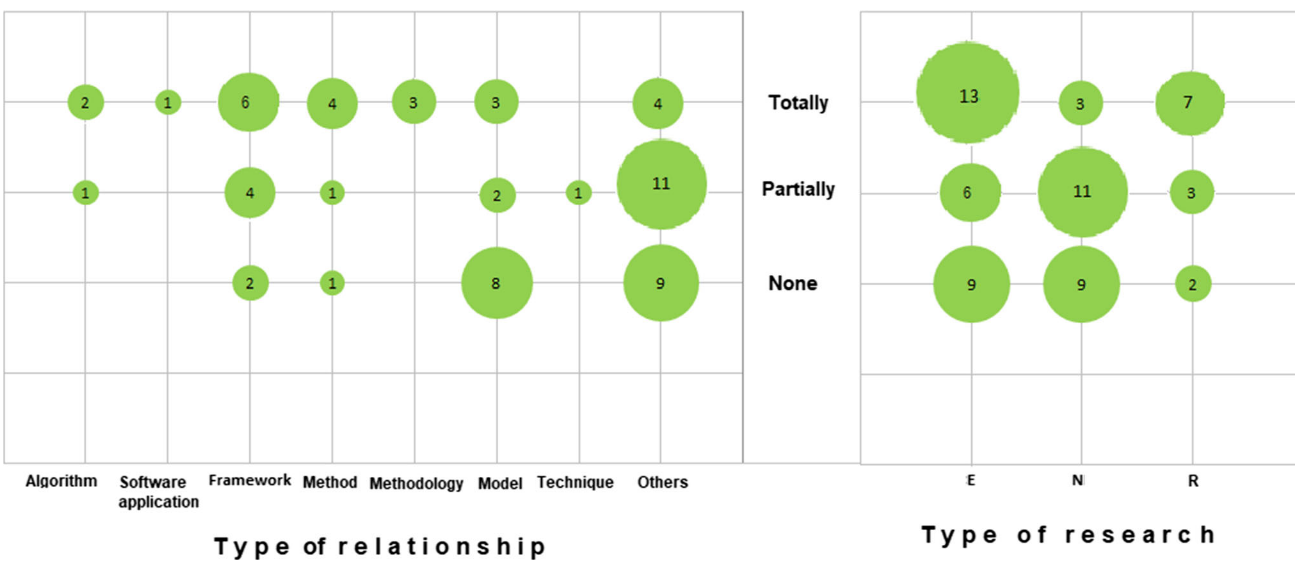


Fig. 9 Systematic review according to three classifications

[66] adds awareness to a reconfigurable circuit within an IoT,[4] device, but for it to gain awareness it must impact all the circuits that make it reconfigurable for this he uses IIT as a referent for awareness.

The concept of awareness provides two platforms for general artificial intelligence (GAI): one for efficient cross-domain optimization and one for machine learning. [10] proposes a framework for testing artificial consciousness and employs IIT as the basis for consciousness. Another topic of interest around artificial consciousness has been the simulation of the cognitive associated with consciousness. In addition, work that has been done in AI and computational science can be used to test theories of consciousness by building models that allow the assumptions of a theory to be made explicit. This type of modeling

[4] Internet of Things.

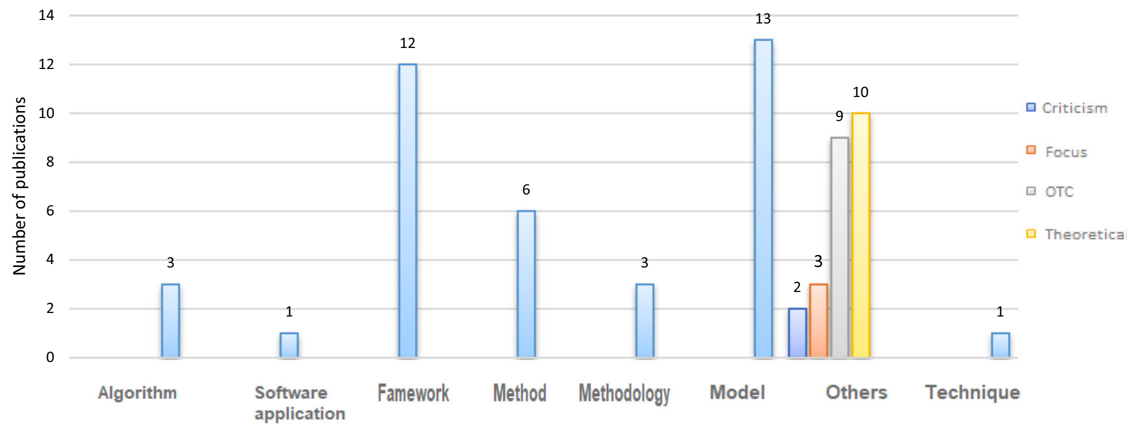


Fig. 10 Number of studies by type of relationship

Table 3 Articles with model as type of relationship

Author	Relationship of use of IIT	Solution
Holland and Gamez [8]	Does not use it explicitly. Theoretical study analyzing IIT and other theories of consciousness	
Wendin [15]	Does not use it. Computational models of neurophysiological models	
Antonopoulos et al. [40]	Uses IIT as a way to measure consciousness in the worm C. elegans	Applicative—Phi measures
Arrabales [41]	Does not use. Capabilities associated with consciousness in artificial systems with cognitive capabilities that can be integrated	
Kralik et al. [22]	Does not use. Defines consciousness as metacognition and in that sense several theories may apply	
Montes [42]	Does not use. Artificial life model that includes functional capabilities of human consciousness in the development of synthetic life. Refers to several theories of consciousness and among them, IIT	
Friedman and Søvik [43]	Uses IIT to measure consciousness in an ant colony. In addition, uses the colony as a way to test theories of consciousness	
Reggia et al. [44]	Does not use. Neuro computational modeling from the perspective of simulated correlates. IIT identified gates as the most conscious components of a neuro controller, based on work by Gamez	
Chandra [23]	Uses IIT. Model that focuses on developing systems that can simulate, recognize, and process human affective experiences referenced to feeling or emotion	
Banerjee and Pal [45]	Does not use it. It is representation for perception-machine encapsulation. The elements analyzed conceptually support IIT and qualia	
Niizato et al. [46]	Uses IIT. Phi can also be applied to general network systems. Applies integrated information in leadership according to degree of autonomy in fish schools	
Marcolli [47]	Uses IIT partially. The idea is to build a different kind of mathematical model based on topology and particular homotopy theory. Interesting to review how topological models would constrain the values of the integrated information	Lower computational complexity based on Friston free energy principle and its computational applications
Javarone et al. [48]	Uses IIT partially. Inspired by work of Tegmark of IIT. Curie–Weiss models are studied. Computational description of Tegmark's consciousness with IIT using Physics and with classical and quantum representation	The 'classical' description of IIT, is achieved through the Ising model

Table 4 Articles with methodology as type of relationship

Author	Relationship of use of IIT	Usages	Solution
Reggia [49]	Uses IIT models of consciousness. Computational modeling is an option in the scientific study of consciousness. Successful models can simulate cognitive, behavioral and neurobiological correlates of conscious processing	IIT as a tool for evaluating theories of artificial consciousness. Monitors and compares information integration with computational experiments instead of measuring Φ	
Iklé et al. [50]	Uses IIT. Experiment to measure the information embedded in a cognitive architecture for two defined tasks	Measures integrated information	Queyranne algorithm for MIP. Python software for Φ* and Phi 3.0 according to Oizumi time series
Popiel et al. [51]	Uses IIT. Attempts to demonstrate that the critical properties of Phi extend to different network connectivities with dynamics governed by the generalized Ising model	Study results give evidence that Phi can capture integrated information in empirical data and according to generalized Ising model	

makes it possible to bridge the gap between abstract theories and the real world and facilitates understanding how consciousness occurs in the brain. Specifically, systems with attributes associated with consciousness such as integrated information have been developed within AI, yet researchers are not confident in attributing phenomenal consciousness to them [8]. To test theories of consciousness [43] present a reverse test based on ant colonies, in order to explore whether scientific theories such as IIT, provide useful and consistent information about the world. Information integration occurs within an ant's brain, and at the colony level emergent forms of cognition arise, it is suggested that ΦMAX occurs at the colony level, representing a possible conscious entity.

In addition, AI has also been used to specify the contents of conscious experience. Cognitive features such as imagination and emotions, perhaps have importance for consciousness and their modeling has been of interest in artificial consciousness. While for [42], it is important to include functional capabilities of human consciousness in the development of synthetic life, in order to capture more possibilities of neurobiological capabilities. Particularly, for the creation of an AGI it is very likely that simulated consciousness will play a very important role [49]. The question whether AI systems can participate in conceptual thinking, seems to be still open [9]. Research on machines with cognitive characteristics, behavior and architecture associated with consciousness are still being studied with much debate in this regard.

Reference [40] refers to ΦE and ΦAR proposed by [87], as a promising way to show relationships between integrated information, consciousness, and other neurocognitive quantities in both real and modeled systems. For [15], cognition and consciousness can be viewed as "emergent global phenomena in a sufficiently biologically complex brain." And as for how cognition can be influenced by quantum physics, one could posit the application of quantum probability theory to human behavior in the decision-making process. Thus, optimization problems must run on quantum hardware to have a quantum advantage.

Now, the brain can be modeled as a recurrent adaptive neural network and perhaps the best that can be achieved is the computational power of these networks. The conclusions around such a hot topic can be reflected in the laws of computation: 1st law you cannot solve non-computable or NP-Hard problems efficiently without a physical infinity or an efficient oracle; 2nd law there are no physical infinities or efficient oracles; 3rd law nature is physical and solves this kind of problems approximately, not efficiently. That is, nature is adaptive and adaptive machine learning algorithms could be used to simulate processes that can find optimal routes. For the time being, networks have been simulated in software because they cannot yet be built with sufficient complexity; The same applies to quantum processing, which can efficiently compute many exponential problems on classical computers, but cannot solve NP-hard problems, so it follows that these problems cannot be efficiently solved by quantum networks [15].

Artificial consciousness deals with models that pose learning. Deep learning can use supervised or unsupervised algorithms. Supervised learning is required in which feedback from the output model updates the input signals and thus minimizes the difference between expected and actual output. Sense organ design would be based on the organs of beings with speech and vision as examples of inputs that can be read, thus, you would have the data to be processed by machines, and the challenge of an artificial consciousness model would be to make sense of that information [1]. Knowing human behavior in information

Table 5 Articles with framework as type of relationship

Author	Relationship of use of IIT	Usages	Solution
Nazri et al. [10]	Uses IIT partially. Framework for testing artificial consciousness by means of a test. Takes IIT as a definition of consciousness with some modifications		
McGreggor [39]	Uses IIT. Establishes a framework that allows to consider experiences as representations of knowledge, and connects theories of consciousness, AI, cognitive science	The concept of conscious experience is taken In IIT, the world experienced by the conscious being is said to be "real"	
Farnsworth [52]	Uses IIT. Quantitative framework that uses IIT as a way to quantify causal independence using IIT measures, and establish autonomy in biological systems	Measuring integrated information	
Kalita et al. [53]	Uses IIT partially. Defines notions of IIT using differential equations among others to measure cause-effect information		
Dodig-Crnkovic and von Haugwitz [54]	Uses IIT. Info-computational framework for modeling artificial and biological cognitive agents. Information is structure for an agent, computation is the dynamics of information	Here, consciousness is taken as a process of information integration	
Tozzi [20]	Does not use it. Presents mappings and projections between different levels of activity according to the multidimensions of the brain		Dynamic Systems—Measuring IIG Python HypeTools manipulates a lot of data
Reggia [55]	Does not use it. Computational mechanisms cannot yet be identified only with conscious cognitive activities. Study consciousness with computational models of biologically based cognitive architectures	Potential computational correlate of consciousness based on system-level properties is the ability to integrate information	
Arrabales [56]	Uses IIT partially. Proposes a set of requirements for what should be an integrative measure of consciousness	Uses elements of IIT within strategies for measuring consciousness in machines	
Aguilera and Di Paolo [57]	Uses IIT. Proposes criticality based on its principles as the susceptibility of a system to changes in its own integration, a property expected to be manifested by cognitive agents	Integrated information as susceptibility to organizational changes close to criticality	
Hunt [58]	Uses IIT partially. Compares IIT with General resonance theory (GRT) and the constellation of qualia of IIT		Suggests a heuristic to compute bounds and resulting capacity
Langer and Ay [59]	Uses IIT. IIT can be applied to neural networks. Combine methods to examine information flows between and within an agent's body, brain and environment	A network-controlled embodied agent in an environment. It has information embedded within the controller, the brain of the agent	
Schneider [60]	Uses IIT. How to determine if an AI is conscious? Three tests are performed	Use IIT as a way to test for artificial consciousness	

processing is important for proper control of machines. An example of a cognitive communication system is presented employing unsupervised deep machine learning techniques and probabilistic generative models. Both deep learning, analytics and data science can have an influence on machine awareness in certain areas, fundamental to artificially recreate areas with sensory information such as vision, perception or artificial speech recognition [23]. Meanwhile, for [54] life processes and, cognition as one of them, fundamentally depend on time. Concurrent computational models are the field that can help us understand the real-time interactive concurrent networked behavior of complex biological systems and their physical structures.

Recent researches in AI and cognitive science have generated proposals on how subjective conscious experience might emerge in embodied agents from models of

Table 6 Articles with method as type of relationship

Author	Relationship of use of IIT	Usages	Solution
Koch and Tononi [61]	Uses IIT. IIT to measure and test awareness	IIT helps to assess whether sick patients, a fetus, etc., are able to have conscious sensations. Consciousness is tested with a practical test	
Arrabales et al. [62]	Does not use it. This work is taken as a method for calculating a guideline score for comparing agents according to consciousness). Opens the possibility to artificial consciousness		
Pizzi and Musumeci [63]	Uses IIT. Method to identify and encode mental states from EEG signals and evaluates their integrated information among others	Depending on the mental states analyzed, the integrated information is evaluated	
Nilsen et al. [64]	Uses IIT. Presents heuristic approaches, methods and applications to find different versions of the integrated information	Specifically employs the measurement of integrated information	Python “Pyphi toolbox”—Matlab “Practical PHI toolbox ”. Heuristic measures
Elamrani and Yampolskiy [65]	Uses IIT. Analytical review of existing tests for machine consciousness	A simple test proposed by [61] to test integrated information	
Liao et al. [17]	Uses IIT partially. Consciousness of a robot is a goal in AI. In IIT experience is basis for consciousness. Cognitive experience can generate consciousness	Integrating information from experiences leads to knowledge and this, in a robot, is the basis for consciousness	

Table 7 Articles with algorithm as type of relationship

Author	Relationship of use of IIT	Usages	Solution
Kanade [66]	Uses IIT. Evolutionary algorithm that uses IIT to measure awareness in machines in their self-repair and configuration process	IIT is used to measure machine consciousness	
Kitazono et al. [67]	Uses IIT. Identify “cores” in brain network or brain areas interacting strongly with each other to understand nature of brain	This paper uses the idea of complex proposed in IIT	Hierarchical partitioning for complex search (HPC algorithm)
Virmani and Nagaraj [68]	Uses IIT partially. New complexity measure for brain networks using a new perturbation-based complexity compression approach	Φ^C measure for brain networks, based on IIT and PCI (perturbative complexity index)	MIB atomic bipartitioning (linear according to network size), instead of MIP.—Matlab ETC ΦC PhiC_ETCFig7.m

Table 8 Articles with technique type of relationship

Author	Relationship with use of IIT	Uses
Barghout [69]	Uses IIT partially. Cognitive computing investigates human information processing in the context of AI and IIT. How languages and images together create knowledge	IIT provides a means to constrain a non-precise perceptual value for the most basic phenomenology

themselves [45, 49] review questions about internal qualia activation, neural correlates of linguistic meanings, cognitive biases in machine learning. The idea is to be able to transpose self-studies to machine embodiment in search of a mind-machine with empathy and within these approaches they conceptually support IIT. [73] recognize the importance of the body in mental development and propose how cognitive skills are based on these capabilities of the body

Table 9 Articles with others as type of relationship

Author	Relationship with use of IIT	Uses	Criticism
Van Hateren [70]	Uses IIT partially. Deals with another Theory of consciousness based on computation, algorithm and neurobiological performance	Takes IIT features: Integration and differentiation	
Manzotti [25]	Does not use it. Shows the theory of consciousness, Spread Min and compares it with IIT		
Shanahan [71]	Does not use it. Study that critically evaluates the anti-functionalist position of some supporters of IIT		Theoretical study critical of the anti-functionalist position of some IIT scientists
Koch and Tononi [61]	Uses IIT. It is a perspective on a momentous issue: whether consciousness will ever be artificially created	Uses IIT to measure consciousness	
Findlay et al. [72]	Uses IIT. Current computers do not have the same phenomenal experience as humans even if there is functional equivalence	With IIT demonstrates that functional equivalence does not imply phenomenal equivalence	
Bedia and Castillo [73]	Does not use it. Presents relationship between cognition and AI and the importance of embodiment, notion of autonomy and agency among others		
Montemayor [74]	Does not use it. The challenges in implementation, formalization and representation presented by IIT require joint work from cognitive sciences, computational sciences and philosophy, among others	Methods of diagnosing disorders of consciousness Foster new insights in information theory and computability New ways of understanding causation Relationship between concept and phenomenal consciousness	Criticism of the implementation, representation and formalization of the theory from a constructive point of view
Patnaik and Kallimani [1]	Although IIT is not used in this article, several uses are presented		
Bach [75]	It does not use it. It reviews challenging problems from AI, promises to the challenges and limitations of such machines in the context of consciousness		Criticism from another theory of consciousness, the CCT
Kelley [76]	Does not use it. Other theory of consciousness and its relations to IIT		
Gauvrit [21]	Uses IIT partially. Another theory of consciousness that takes elements from IIT		
List [77]	Does not use it. Algorithmic approach to cognition. IIT connects consciousness to information theory	IIT to support that group agents have little or no collective phenomenal consciousness	
Velmans and Schneider [78]	Uses IIT. Consciousness in groups or collectives		
Thagard and Stewart [79]	Does not use it. Foundation on consciousness		Raises criticisms from another theory of consciousness (SPC)
Calvo et al. [80]	Does not use it. Shows fundamental aspects of consciousness from the point of view of SPC theory and IIT	Considers consciousness in plants in meristems and uses integrated information	
Langer and Ay [59]	Uses IIT. With evolutionary approach reports evidence that indicates that meristems in plants act consciously perhaps at the level of minimal consciousness		

Table 9 (continued)

Author	Relationship with use of IIT	Uses	Criticism
Kleiner [81]	Uses IIT partially. Proposes a measure, Causal Information Integration CII, that measures intrinsic causal cross-influences in environment with unknown external influence, complies with [6]		Raises criticisms of the presented theories of consciousness
Baudot [82]		Presents aspects in which it reconciles IIT with GWT	
Safron [83]	Uses IIT partially. Mathematical review of various theories of consciousness on the argument that these cannot explain consciousness	Combines theories such as IIT and GNWT in a unified way	Integrated information implies consciousness only for certain systems
Stephan and Klima [9]	Uses IIT partially. Review presenting well-established results, compares the resulting theory with IIT among other theories		
Mallatt [84]	Uses IIT partially. Integrated World Modeling Theory (IWMT) uses the free energy principle and the active inference framework (FEP-AI), and	Evaluates IIT	
BarthelmeÃ [85]	Uses IIT partially. Whether AI systems engage in conceptual thinking, as opposed to perceptual thinking seems to be an open question	Uses IIT and GWT to apply them to a reasoning system in a common sense and cognitive reasoning scenario	
Moruzzi [86]	Uses IIT partially. Compares IIT with a theory derived from neurobiological naturalism, consciousness is an evolutionary feature of complex brains	Integrated information that a system for generating music that uses GNWT has and what this says about its creativity levels	

both in embodiment and situationality and kinesthesias. This has led to currents related to the classical computational paradigm, as well as to others such as inactivism, which study those aspects ignored by computationalism (agency, autonomy, etc.). The IIT formalism provides an explanation of the causal structure of a system and its integrated information measures the causal constraints that the system exerts on itself and can peak at a macro level of description [88, 89]. The causal principles of IIT can also be employed to identify and quantify the actual causes of events ("what caused what"), such as the actions of an agent. For this, a simulated agent equipped with a small neural network is used [90].

This review also considered the development or definition of models that take IIT into account. [49] presents a review of approaches or theories of consciousness, GWT, IIT among others, as key concepts in the creation of computational models of consciousness, mainly used in artificial consciousness and that could be complementary in their study. For [44] "recognizing the difference between simulated correlates of consciousness and instantiated machine consciousness clarifies the nature of the progress that has been made in artificial consciousness research over the last two decades"; although phenomenal consciousness, let alone evidence of it, has not yet been shown in a machine.

RQ2.2 ¿What applications exist that use or refer to IIT?

To answer this question, the publications were classified as total, partial or none, depending on whether there was any use of the theory (total), some of its characteristics were used (partial), or there was no explicit use of the theory (none). Table 10 presents a classification of the uses made of IIT.

To answer this question, the publications were classified as total, partial or none, depending on whether there was any use of the theory (total), some of its characteristics

Table 10 Classification according to the different uses of the IIT

Usage	Description
measureC	Measure consciousness
measureAC	Measure artificial consciousness
TestC	Test for consciousness
TestAC	Test for artificial consciousness
II	Integrated information
measureII	Measure integrated information
Case	Case example of application
CE	Conscious experience
Components	Basis

were used (partial), or there was no explicit use of the theory (none). Table 10 presents a classification of the uses made of the IIT.

It was found that 24 of the 63 publications presents uses of the IIT. Figure 11 shows the ways in which the theory is used according to the type of relationship, and Table 11 discriminates the publications that present such uses:

Clearly, IIT is associated with consciousness, but the concept of integrated information has proven useful in other respects. For example, to quantify the "totality" of causal networks in the case of systems with circular causality, fundamental to the emergence of a system's autonomy or degree of self-determination [52]. Studies suggest that integrated information could be an order parameter in complex systems, similar to the generalized Ising model. Complex biological systems are irreducible because of their intrinsic causal structures, and by using IIT in these, Phi could be considered as a measure of the totality of an autonomous system and viewed as its order parameter. [90] use a simulated agent equipped with a small neural network to demonstrate that the causal principles of IIT can also be employed to identify and quantify the actual causes of events ("what caused what"), such as the actions of an agent. Employing neural networks, [59] combine different methods to examine the information flows between and within the body, brain, and environment of an agent controlled by a neural network and its integrated information.

On the other hand, as an initial effort to model collective animal behavior, IIT has served to interpret classifications of cellular automata, animats, Boolean networks, etc., [46]. In Reference [57], see integrated information as susceptibility to organizational changes close to criticality.

From the perspective of artificial consciousness [66] opens the way for hardware self-repair and self-configuration without human intervention by employing an evolutionary algorithm inspired by "Darwin's theory of evolution" and the IIT, used to measure the level of consciousness of a reconfigurable circuit within an IoT device. The results obtained by calculating the value of Φ support the proposed concept of machines with some degree of consciousness. Now, as [44] point out, having simulated correlates of consciousness as a focus, neuro computational modeling has been successful for machine simulations, such as finding that IIT identifies activation modules as the most conscious components of a neuro controller [91], linking activation mechanisms in cognitive control with studies of consciousness [92].

Another area with applications of IIT concerns tests on artificial consciousness. [43] uses the ant colony as a reverse test for theories of consciousness such as IIT, which could be useful for analyzing collective consciousness and behavior using scales in natural and artificial worlds. [65] review contemporary tests relevant to examining artificial consciousness, such as:—Tests according to Φ based on IIT;—[61] test for measuring Φ, which requires a machine to differentiate key features in a scene from many possible scenes;—"the six key factors of human-like consciousness", a test designed by Goerztel that unifies global workspace theory and IIT as necessary manifestations of expertise. Furthermore, the IIT could be considered as an objective and quantitative tool for evaluating theories of artificial consciousness [49].

Reference [50] presents another way of using IIT in which Phi could be seen as a measure of a consciousness-related property (holistic integration of information) of a complex cognitive system, and some of the authors lean toward a nuanced view in which Phi is an estimator of one among many properties of consciousness in human-like cognitive systems. However, the practical estimation of Phi in cognitive systems has value independent of debates about the fundamental interpretation.

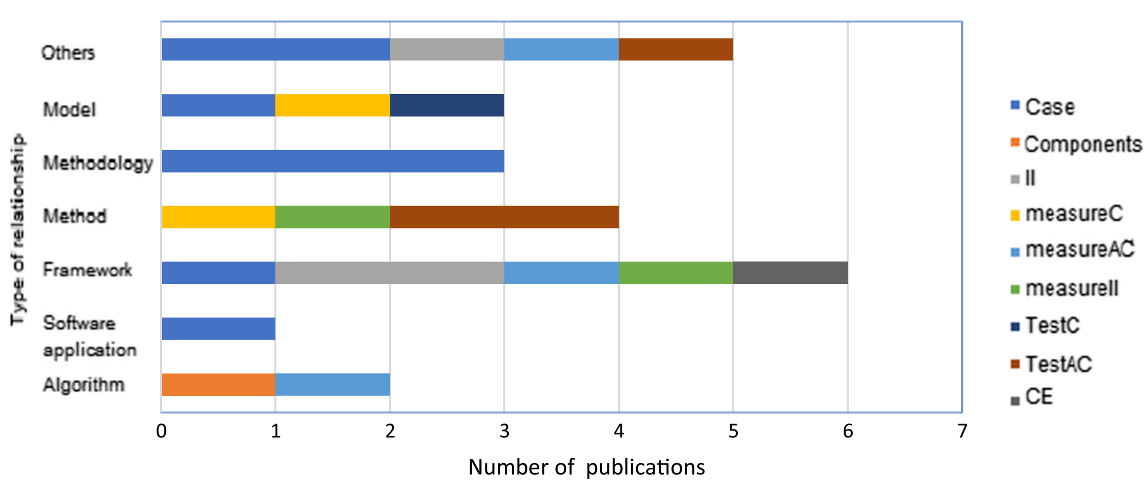


Fig. 11 Uses according to type of relationship

Table 11 Publications with at least one use of ITI

Authors	Type of publication	Type of investigation	Type of relationship	Use
Kanade [66]	Book section	E	Algorithm	measureAC
Koch and Tononi [61]	Journal article	R	Method	TestAC
McGreggor [39]	Book section	R	Framework	CE
Koch and Tononi [26]	Journal article	R	Others	TestCA
Pizzi and Musumeci [63]	Journal article	E	Method	measureC
Antonopoulos et al. [40]	Journal article	E	Model	measureC
Findlay et al. [72]	Journal article	E	Others	measureAC
Nilsen et al. [64]	Journal article	E	Method	measureII
Farnsworth [52]	Journal article	N	Framework	measureII
Dodig-Crnkovic and von Haugwitz [54]	Book section	R	Framework	II
Elamrani and Yampolskiy [65]	Journal article	E	Method	TestAC
Friedman and Søvik [43]	Journal article	E	Model	TestC
Reggia [49]	Journal article	N	Model	Case
Iklé et al. [50]	Conference proceedings	E	Methodology	Case
List [77]	Journal article	R	Others	II
Calvo et al. [80]	Journal article	E	Others	Case
Aguilera and Di Paolo [57]	Journal article	N	Framework	II
Kitazono et al. [67]	Journal article	E	Algorithm	Components
Albantakis [90]	Journal article	E	SW application	Case
Niizato et al. [46]	Journal article	E	Model	Case
Langer and Ay [59]	Journal article	R	Framework	Case
Popiel et al. [51]	Journal article	E	Methodology	Case
Schneider [60]	Book section	R	Framework	MeasureAC
Moruzzi [86]	Conference proceedings	R	Others	Case

Reference [77] studies the possibility of existence of group agent consciousness (collective consciousness) using integrated information for that purpose. [43] shows how a level of Φmax higher than that of individual aggregated individual worker ants could occur in the ant colony, representing the colony as a conscious entity.

From an approach grounded in the scientific study of consciousness, some support can be found for complicated cases, such as coma patients or human-level AI, and a broader perspective on what we mean by "conscious" could also be given [71]. In the future, IIT could transform methods for diagnosing disorders of consciousness and improve the treatment of critically ill patients [74].

On the other hand, [40] uses ΦAR as a useful measure of the information contained in the dynamic brain network of the soil worm C. elegans.

Broader looks at applications of IIT are in [74], for whom IIT should be studied beyond consciousness, as it could bring new insights into information theory and computability by differentiating integrated information and would provide other ways to interpret causality and thus impact research in many areas.

RQ2.3 ¿What solutions are presented regarding the challenges posed for the development and/or application of IIT?

Ten publications were found that present some type of solution for the materialization of the IIT. In Tables 3, 4, 5, 6, 7, these publications, in the solution column, present information on the type of solution found.

Reference [93] tried to monitor the integrated information over time by doing computational experiments instead of calculating Φ of the state-by-state form. It was found that the effective integrated information is maximized for a network with strong distributed connectivity. Reference [27] computed Φ systematically for 4-neuron networks and compared it with the "liveness" or liveliness of the network and showed that these two measures are strongly correlated in such small networks [49].[5] Reference [87] shows some practical methods to measure integrated information from time series data, for which they propose two new measures ΦAR and ΦE and that can be calculated using time series data that can come from model or real

[5] Another measure of network interactions as defined by [93].

systems. ΦAR is well suited for cases where the time series does not have Gaussian distribution, but stationary and stochastic. By construction, when applied to stationary data with Gaussian distribution, it is equivalent to the empirical version of phi, ΦE. It should be noted that these measures already differ when applied to stationary non-Gaussian data. In fact, "ΦAR provides a useful measure of integrated information based on the relationships between conditional entropy, partial covariance, and linear regression prediction error." The authors indicate that they calculated Φ by employing a macroscopic partitioning of the network. For the human brain case, it has not yet been proven that there is correlation between the results of Φ at the macro level with those of Φ at the micro level. Reference [40] use ΦAR, a measure of auto-regressive integrated information to possibly reflect degrees of consciousness in the C. elegans worm brain network. For the estimation of Φ, they employ the Matlab code "ARphidata.m" for stationary data provided by [87].

Reference [59] proposes ΦCII, another measure of integrated information and use an iterative information geometric algorithm, the "em-algorithm". This algorithm will converge to the minimum, although it could be a local minimum, so the algorithm should be run several times to find a global minimum. Using this algorithm, one can compare the behavior of ΦCII with existing integrated information measures.

Another approach is that of [58] from the General resonance theory (GRT) consciousness theory developed by Hunt [58] himself in the company of Schooler. Hunt [58] describes how GRT compares with IIT and works on a heuristic to compute the bounds and resulting capacity for phenomenal consciousness in resonant structures. The IIT constellation-qualia characterization framework is compatible with GRT and can be a useful tool in conjunction with the GRT quantization framework.

A possible approach to reduce computational complexity in the computation of integrated information can be derived from [94] free energy principle and its informative applications [47].

The work by [64] is based on state-independent heuristics and approximations, so they cannot be directly compared to Φ3.0 (state-dependent). Some of these proposals can be applied on large systems and also contrasted with Φ. Although some of them have underlying aspects of the IIT and others have been taken to practical applications, it has not been possible to verify that they really quantify the integrated information according to the IIT3.0 and therefore, if through them the theory can be tested. Heuristics have been applied to electrophysiological data, time series of continuous variables and discrete variables and little has been tested with respect to Φ3.0. The authors propose to calculate a quantity, Φpeak3.0, independent of state that corresponds to the maximum value of Φ3.0 for all network states and will allow comparisons to be made with the heuristic and approximate measures reviewed. This new measure is more a measure of the network's capacity for awareness than its currently achieved level of awareness. Alternatively, the mean value of Φ3.0, which has some relation to the state-dependent Φ3.0 value, was also calculated under certain regularity conditions with very similar results. A review based on the measures proposed in [64] is presented in Table 12, this information can be expanded in "Appendix 2". Measures for integrated information.

Reference [67] introduces the "Practical PHI toolbox for integrated information analysis" to implement measures of integrated information with formulas for discrete shapes, employing an exhaustive search algorithm to find the MIP for bipartitions (MIB) and a normalization factor according to IIT2.0. Heuristics were tested for those state-independent measures with time series data generated for the entire network [64]. Reference [68] develop "Phi-C_ETCFig7.m", a Matlab program that computes ETC ΦC, another measure of integrated information and address the practical problem of MIB with atomic bipartitions, which increases linearly with lattice size. Other researchers have also recommended atomic bipartitions instead of MIP.

Reference [50] measured the integrated information of a cognitive architecture (OpenCog) for two specific tasks;[6] they took as reference a tool in Matlab to estimate Phi, from coupled time series, designed by [101], which integrates Queyranne's algorithm and Φ computation to find the MIP. A novel methodology was adopted, first applying Independent Component Analysis (ICA) to reduce the original set of sparse time series to fewer dense time series (reduce dimensionality of the problem) and then calculating Φ, from the dimensionally reduced time series; the sum of squares of residuals for each dimension was also used by selecting it with the minimum total SSR. ICA and Phi have common support in the mathematics of mutual information that provides some consilience. In addition, in the experiment conducted with the humanoid robot Sophia, two methods were presented to calculate Φ: One based on the calculation of Φ* (an approximate measure of Φ) and the other is the calculation of Φ3.0, both introduced by Oizumi. For their solution they took Φ3.0 and the calculation of probability distributions from [102] as a reference, it was developed in Python and based on the Matlab solution of [101], and to store the probability distributions Python dictionaries were used instead of arrays.

Reference [67] proposes the fast HPC (Hierarchical Partitioning for Complex Search) algorithm that searches

[6] reading short documents and guiding the humanoid robot Sofia in performing a dialogue-based interaction.

Table 12 Some measures for integrated information

State-dependent measure	State-independent measure	Characterístics	Source
$\Phi_{3.0}$	$\Phi_{3.0}{}^{peak}$	Integrated information (II) according to $IIT_{3.0}$	Oizumi et al. [6]
CO $\Phi_{3.0}$	CO $\Phi_{3.0}{}^{peak}$	"Cut-one" Cut a connection when partitioning. Reduces the scope of the MIP search on possible system cuts	Mayner et al. [95]
$NN\Phi_{3.0}$	$NN\Phi_{3.0}{}^{peak}$	"No new concepts" No new concepts after partitioning. It is possible that the CES of the partitioned system has reducible concepts in the non-partitioned system	Mayner et al. [95]
$WS\Phi_{3.0}$	$WS\Phi_{3.0}{}^{peak}$	System as a whole. All nodes in the system are included in the CM	Nilsen et al. [64]
$IC\Phi_{3.0}$	$IC\Phi_{3.0}{}^{peak}$	Iterative cuts. Included in the larger complex is the subsystem where nodes with nonrecursive connectivity or an unreachable state are iteratively removed	Nilsen et al. [64]
	$Est.n\Phi_{3.0}{}^{peak}$	Estimation of $\Phi_{3.0}$ peak from n states ($n \leq 15$). Take a few states instead of taking the maximum over all possible states	Nilsen et al. [64]
$\Phi_{2.0}$	$\Phi_{.2.0}{}^{peak}$	Integrated information according to $IIT_{2.0}$	Balduzzi and Tononi [27]
$\Phi_{2.5}$	$\Phi_{.2.5}{}^{peak}$	A combination between features from $\Phi_{2.0}$ and $\Phi_{3.0}$	Tegmark [96]
	D1	Achievable states. The number of states that were reachable is calculated, quantifying repertoire of available states of the system	Marshall et al. [97]
	D2	Cumulative variance of elements of the activity of the system nodes given the maximum entropy distribution of the initial conditions	Marshall et al. [97]
	S	Entropy of coalition sample. Entropy measure of the distribution of observed states indicating average diversity of a system of visited states	Schartner et al. [98]
	LZ	Functional complexity. Signal complexity measured by algorithmic compressibility through Lempel–Ziv compression	Schartner et al. [98]
	Φ^*	Integrated decoder-based information on top of $IIT_{2.0}$	Oizumi et al. [99]
	SI	Integrated stochastic interaction. Based on $IIT_{2.0}$	Barrett and Seth [87]
	MI	Mutual information. Based on $IIT_{1.0}$	Oizumi et al. [100]
	Geometrical integrated information	Derived measure of integrated information [100] with statistically disconnected causal influences, which is interpreted geometrically as the divergence between the current and approximate probability distributions of the system	Tozzi [20]

Adapted from [64]

complexes by hierarchically partitioning subsystems, as an effort to find the integrated information in a better way.

According to [77] from heuristics it could be inferred which systems tend to present high Φ values and which tend to low Φ values. Reference [28] shows that high values of Φ tend to occur in systems with many internal feedback cycles, while low values of Φ tend to appear in direct-feed systems. "In a pure direct-feed system, one layer feeds the next without recurrent connections. The input layer is completely determined by the external inputs and the output layer does not affect the rest of the system."

Reference [103] introduces a model in which they modify the IIT framework and introduces some simplifications and approximations that allow measuring the integrated information when the system size grows considerably. For this, the authors use kinetic Ising model, which proposes a way to adapt the IIT measures to really large systems. To find the MIP, the connectivity of the homogeneous system is exploited for nodes in the same region. Thus, for systems of infinite size, the MIP will be one of the possible partitions that isolates a node of the system. It is assumed that under certain conditions the MIP is either a partition that cuts a node from the system mechanism or a cut that separates entire regions into distinct partitions reducing the cost of computing the integrated information.

Reference [104] establishes a correlation between the integrated information and the topological dimensionality of the attractor dynamics continues to be shared. By using "delayed inclusion techniques" one can evaluate effects of unobserved nodes on attractor dynamics, capable of quantifying the dimensionality of an embedded attractor even in partial observations. According to [20], topological approaches also allow the extension of current notions of IIT to continuous dynamical systems. Tozzi [20] employs a measure of geometrically interpreted embedded information and uses "HypeTools", a Python toolbox for the visualization and manipulation of large, high-dimensional datasets.

RQ6 ¿What criticisms or problems are raised about IIT?

Reference [20] raises a problem with the measurement of integrated information from neural signals that necessitates the evaluation of all elements at the same time; moreover, the interpretation of the spatial partitioning requested is quite confusing when dealing with continuous time series variables obtained from nonlinear dynamics.

Several mathematical problems arise when faced with the quantification of multiple influences, such as overestimation and confounding non-causal influences. To provide an answer to this [99] generated a measure of integrated information with statistically disconnected causal relationships between elements, thus integrated information can be understood geometrically as the divergence between the actual probability distribution of a system and an approximation. This achieves a methodological unification of a large number of information theoretic measures, such as stochastic interaction, mutual information, transfer entropy etc. "Integrated geometric information" allows "quantifying the strength of multiple causal influences between elements, projecting the probability distribution of complex systems in constrained multidimensional garbage cans" [20].

Reference [105] presents The Cortical Conductor Theory of consciousness, and in his exposition, he raises criticisms of the foundation and substantiation of IIT. Likewise, [79] raises criticisms of IIT but from the Semantic Pointer Competition (SPC) theory of consciousness. From a constructive point of view, Reference [74] presents a critique of the implementation, representation and formalization of the theory.

Reference [71] raises a critique of the anti-functionalist position of some scientists inspiring IIT.

5 Future work and challenges

The analysis of the results shows that IIT is a growing field of research not only in neuroscience but also in areas such as artificial intelligence and cognitive science. Because of this, there are research opportunities in analysis, modeling, measurement and testing for integrated information.

Regarding machine cognition, there are several challenges such as automation of learning from external environments without human intervention and adaptive learning. It is posited that cognition should occur from feedbacks from external factors or other peer machines and application of machine learning to process the large amounts of data [1].

The challenges in implementation, formalization and representation presented by the IIT framework require joint work from disciplines such as cognitive sciences, computational sciences and philosophy among others [74].

RQ4 ¿What trends are there regarding the development of IIT and what future work is on the horizon?

How to algorithmically account for the same mental biases with more or less computational power? It is not yet known exactly what would be required for this. More research should be done in determining whether theorems such as the algorithmic encoding theorem are valid under super-universality or sub-universality conditions [21].

On the other hand, work on neural correlates of consciousness (NCC) will lead to substantial progress in neuro phenomenology and thus predictions about artificial consciousness [8], moreover, they are used clinically to treat disorders of consciousness and develop methods for two-way communication, training, and control [15]. Much work in artificial consciousness would benefit from characterizing phenomenal state implementations [62]. Fully characterizing high-level cognition through a neuro-computational representation could provide necessary and sufficient evidence for artificial phenomenal consciousness that would benefit from possible computational correlates of consciousness [55]. Similarly, [54] raise the importance of advancing on the connections between high- and low-level cognitive processes, as well as finding the relationships between cognition and consciousness.

Of great importance are computational models not only to test assumptions about consciousness, with experimental data but, for example, to reveal implications of research hypotheses that result in impractical behaviors for neuroscience and cognitive science. Models are a way to arrive at machines with behaviors associated with human consciousness [44, 49]. Recurrent neural networks using machine learning systems have been successful in discovering and predicting the shape of visual and auditory stimuli, while deep networks can help with hierarchical representations of knowledge [105].

Identifying consciousness through behavioral interpretation remains an open problem. However, measuring the intrinsic property of the concept of consciousness requires more effort [10]. Reference [62] suggest the design of cognitive tests for restricted problem domains, and [63] in their studies on mental states by analyzing brain dynamics to quantify and evaluate their integrated information, are experimenting with a new set of underlying visual, auditory and cognitive stimuli, as well as comparing them with emotional stimuli.

With the advancement of quantum information theory, it would be pertinent to review how IIT can benefit in this regard and, how it would relate to current theories regarding the quantum nature of the conscious mind [49]. Likewise, for [54] in relation to quantum mechanics, information processes represent laws of physics, and a field of research emerges in the exploration of the space of natural computation and the relationships between information, computations and the physical.

On the other hand, [106] has studied the criticality and behavior of Φ in a classical Ising model extending the concept of criticality to IIT. Reference [51] posits that the criticality properties of Φ extend to networks of connectivity with dynamics according to the generalized Ising model. This model has been shown to simulate the statistical behavior of the brain. By generating Φ for random networks, a methodology is outlined that can be applied with patient tractographies to create a new field of development for IIT.

Another aspect to consider are the links between empirical research in cognition, neuroscience etc., and theories of consciousness that could bring new predictions by establishing better relationships between computational models and experimental data [49].

Referring to the computation of Φ, [64] proposes to take a compositional approach by combining values from subsets of measures, preferably than using all measures at once. Another thing is that the heuristics do not correlate with Φ2.0 which is state dependent. The directionality of the cut is also considered, since Φ3.0 uses unidirectional cuts (separates a directed connection) however other heuristics use bidirectional cuts (Φ2.0, Φ2.5) and also total cuts are used that separate system elements (Φ*, SI, MI), leading to an overestimation of the integrated information. Another aspect to review concerns the data used in the different measurements, e.g., the full TPM was used for the approximate measurements, Φ2.0, Φ2.5, D1, D2, the other heuristics in Table 3 were calculated based on the generated time series. Both time series and TPM can describe deterministic networks since the system was initialized to all possible states at least once, but data from deterministic systems might not be sufficient from time series and must be perturbed regularly. This could be solved by adding noise to the system thus avoiding fixed points, but, in addition, the work could be oriented to control the number of samples and check the impact of non-self-sustaining activity as many of these revised heuristics depended on the size of the generated time series. Table 3 shows some proposed measures and approximations based on [64] for which it has not been shown that they actually quantify the integrated information according to IIT3.0, and thus can be used to test the theory.

RQ5¿What barriers are encountered for the advancement of IIT?

The proposed algorithms for computing integrated information have multiple factorial dependencies and for a network all partitions of possible subsets are needed. An analysis of a network of 18.000 nodes with the first version of the integrated information [107] would take 109.000 years, and the times for the measure proposed in [27] are also too considerable. In addition, greater precision is needed for the current measures of integrated information because different values can be obtained depending on the algorithm used. According to [49], IIT requires further theoretical development and thanks to its conceptual framework and quantitative methods, it would be useful to review whether possible new insights can lead to verifiable predictions with computational models.

The problem of testing consciousness still remains open in biological organisms; however, biological approaches are currently being analyzed to determine its applications in artificial agents [62]. An additional problem is that there are no clear specifications of behaviors associated with consciousness. Functional specifications for artificial consciousness could materialize in computational code [8]

Reference [44] defines the computational explanatory gap as the lack of understanding of how high-level cognitive information processing can be mapped into low-level neural computations, which in turn hinders the creation of instantiated artificial consciousness. The gap cannot be understood as a problem specific to computers or computational science, but rather a broad form that refers to how one class of high-level computations can be mapped into a fundamentally different type of low-level computations, regardless of the hardware used. Likewise, [25] posits how the gap between semantics and syntax in information processing is a problem for AI since one cannot appeal to emergent properties hidden in the brain.

One of the most important drawbacks of many current machine learning paradigms is the particular focus limiting itself to classification and rule learning. Our minds are not classifiers but simulators and experimenters, as with machine learning, the mind learns to distinguish particularities in sensory input, and then combines them, in a more complex way, and organizes them into maps. Integration of functions can occur in geometries and dynamic objects, patterns and motor procedures, etc. [105]. Reference [93] presents a "study that has used attractor neural networks as associative memories that store images, based on unweighted neurons, to review how different connection patterns and architectures influence the integration of estimated information from a network". For this from computational experiments, the integrated information is monitored over time, and thus avoids finding ϕ for each state. Moreover, it could be seen that the effective integrated information was maximized in a network with strong distributed connectivity.[7]

A major limitation for IIT is the practical barriers to measure ϕ in any network (except small networks) [55]. Sample of this is the computational cost of the current version, Φ3.0, which grows in order O(n53n) [64]. IIT3.0 is more computationally intractable than previous versions, in fact Reference [77] relies on Aaronson who surmises that computing Phi, is a computationally hard problem.

A major challenge for IIT is to find the MIP [50]. While to compute Φ3.0 one needs to find a single MIP at the system level, it is also true that, one MIP must be found at the mechanism level for each possible combination of mechanism and purview, implying exponential growth as the system size increases. Approximations and heuristics such as the cut-one approximation can be revisited [64].

Another difficulty about the dynamic IIT approach is that complete specifications of the underlying dynamics of the real signals and how to identify them are not yet available, because of the lack of a strong quantization method [63].

Finally, [55] mentions that even though computational modeling has become accepted as a tool for the study of consciousness, there is still no computational approach to artificial consciousness that has presented some evidence for phenomenal consciousness in a machine, or that is eventually possible. In this regard, [61] considers that a different kind of machine can be thought of, in which knowledge of the relationships of the world we live in can be incorporated into a single, highly integrated system, and quite possibly to achieve such high levels of integration, this machine could take advantage of the foundations underlying the mammalian brain.

6 Conclusions

It is necessary to develop computational tools with sufficient power and algorithms that take advantage of it to study the complex brain system, since this is an area that offers many possibilities for study. Consequently, the incorporation of technologies such as IoT, semantic web, cognitive computing and machine learning could be oriented toward the simulation of aspects of machine consciousness. Research on the development of artificial personality in artificial consciousness could also be presented. Similarly, advances in self-awareness and the human brain have been aided by deep learning adaptive algorithmic strategies coupled with computational power and advances in the study of dynamic brain imaging.

The possibility of providing tractable polynomial-order solutions to NP-hard problems from unconventional classical computing is still being actively explored. This kind of computing is a rather broad field that includes topics such as quantum computing (QCOMP), biocomputing, optical computing, analog computing, hypercomputing, chemical computing, molecular computing, among others. The idea that unconventional computing paradigms (UCOMP) can provide better solutions than those provided by classical or digital computing is becoming increasingly common. Another novel computing paradigm is Memcomputing, which employs non-local dynamic systems in time to perform memory-based computations. The digital version of these machines (digital memcomputing machines, DMM) is a possibility to solve combinatorial optimization problems, and ordinary differential equations could be used to describe the circuits. Tests have been done that have shown linear DMM simulations in time and memory on very large problems compared to the results of current exponential solutions.

The main objective of this research was to identify and examine the state of the art on integrated information theory and its relationships with artificial intelligence and cognitive sciences to find answers to the challenges posed by the theory in terms of integrated information measurement. It is noteworthy that while upper bound estimates of Φ3.0 have been proposed for some system size, there is little information on the actual distribution of Φ3.0 in networks with different types and topologies. For example, in studies with attractor neural networks it could be seen that the effective integrated information became maximal in a network with strong distributed connectivity. It is also not very clear what kind of dependence there is between the underlying graph topology and its associated dynamics, as the dynamics and observed functionality are largely determined by the topology of the graphs. This opens a great field of research especially looking for the relationship between the substrate topology and the informational field. It would be highly desirable to investigate on the relationship between Φ3.0 and network properties such as weight, noise, thresholds, element types among others. These relationships could have implications in determining the properties of biological systems for high Φ3.0, as well as properties that facilitate the production of consciousness in artificial systems.

Future research in IIT requires algorithmic efficiency to approximate Φ values in large neural networks and that these can be applied to large-scale models of the brain to make comparisons with values obtained from other

[7] Exhibiting some relationship to the global workspace awareness theory (GWT).

networks, for while theoretically Φ is applicable to any physical system, currently in practice it is not computable, as it only applies to systems with approximately 12 nodes. These practical barriers to measuring φ are a major limitation, as in the case of the current version Φ3.0, making it more computationally intractable than previous versions. Moreover, this is probably one of the reasons for the relatively little computational work in this regard. However, to overcome the barriers generated by the computational costs, it is recommended on the one hand to implement approximations to reduce these costs, and on the other hand, to use heuristic measures that capture the essence of the theory to try to materialize them with more tractable methods. However, in studies conducted, approximate measures performed well with respect to predicting the value of Φ3.0 although they were computationally intensive and required full knowledge of the system's TPM, while heuristic measures showed reductions in time and knowledge required, but did not approach Φ3.0. Because the solutions found are functional for very small systems, the big challenge for IIT is to find the MIP.

In particular, the cut-one approach could provide good computational performances. Consequently, an extension of this approximation or some variation of it, from system-level MIP to mechanism-level MIP, is suggested in order to achieve better computational efficiency. Another suggestion in this regard has to do with the possibility of exploring partitioning schemes other than the bipartitions considered by IIT3.0 when searching for MIP, for which it is suggested to review different rules for the cuts allowed as potential approximations. In line with what has been previously described, if certain considerations are taken into account, it would be interesting to determine the classes of networks where the heuristics tested are in tune with Φ3.0.

Another gap in the current literature is the lack of direct comparisons with Φ3.0, since due to the computational cost already mentioned, it is not possible to validate the proposed measures on networks of interest, however, as a way to overcome this gap one could take smaller systems and validate the measures on them, since Φ3.0 could be computed directly. One could think of evaluating the accuracy of approximations and heuristic measures of integrated information with respect to Φ3.0, with deterministic, isolated, discrete networks of binary logic gates such as those of IIT3.0.

Finally, whether the concept of integrated information can eventually explain consciousness or not, it may be useful in describing the properties necessary for autonomous organisms. Furthermore, regardless of the inference of IIT for consciousness, integrated information is taken as a quantitative concept useful in describing the properties necessary for autonomous agency and with the possibility of many applications. Consequently, it is of utmost importance to consider the generation of models that support the computation of integrated information and in particular the finding of MIP, supported by mathematics and artificial intelligence with the use of heuristic and/or approximate techniques to improve the current solutions.

Appendix 1

Id	Title	Authors
P1	A theory of consciousness computation, algorithm, and neurobiological realization	Johannes Hendrik van Hateren
P2	A Hybrid Evolutionary Algorithm for Evolving a Conscious Machine	Vijay A. Kanade
P3	A New Theoretical Framework for Testing Consciousness in a Machine	Azree Nazri, Abdul Azim Abd Ghani, Izuan Hafez, Keng-Yap Ng
P4	A Physicalist causally oriented Foundation for a Conscious Machine based on the Spread Mind	Riccardo Manzotti
P5	A test for consciousness	Christof Koch, Giulio Tononi
P6	ConsScale A Pragmatic Scale for Measuring the Level of Consciousness in Artificial Agents	Raul Arrabales M., Agapito Ledezma E., Araceli Sanchis de Miguel
P7	An Experience Is a Knowledge Representation	Keith McGreggor
P8	Artificial Intelligence and Consciousness	David Gamez, Owen Holland
P9	Ascribing Consciousness to Artificial Intelligence	Murray Shanahan
P10	Can Biological Quantum Networks Solve NP-Hard Problems?	Goran Wendin

Id	Title	Authors
P11	Can machines be conscious	Christof Koch,Giulio Tononi
P12	Coding Mental States from EEG Signals and Evaluating their Information Content: a Computational Intelligence Approach Integrated	Rita Pizzi, Marialessia Musumeci
P13	Dynamical complexity in the C.elegans neural network	Chris G. Antonopoulos,Athanasios S. Fokas; Tassos Bountis
P14	Dissociating Intelligence from Consciousness in Artificial Systems – Implications of Integrated Information Theory	Graham Findlay,William Marshall, Larissa Albantakis, William Mayner, Christof Koch, Giulio Tononi
P15	Evaluating approximations and heuristic measures of integrated information	André Sevenius Nilsen; Bjørn Erik Juel; William Marshall
P16	Evaluation and development of consciousness in artificial cognitive systems	Raúl Arrabales M
P17	Hacia una teoría de la mente corporizada: la influencia de los mecanismos sensomotores en el desarrollo de la cognición	Manuel G. Bedia, Luis Fernando Castillo Ossa
P18	How organisms gained causal independence and how it might be quantified	Keith Douglas Farnsworth
P19	Implementation, formalization, and representation: Challenges for integrated information theory	Carlos Montemayor, J. Acacio de Barros, Leonardo Guimaraes de Assis
P20	Informational structures and informational fields as a prototype for the description	Piotr Kalita, Jose A. Langa, Fernando Soler-Toscano
P21	Hypernym and Spatial-Taxon Hierarchy. A Cognitive Informatics Fuzzy Logic Approach to Combining Linguistic and Image Taxonomies	Lauren Barghout
P22	Metacognition for a Common Model of Cognition	Jerald D. Kralik, Jee Hang Lee, Paul Rosenbloom, Philip C. Jackson, Oscar J. Romero, Samuel Epstein, Ricardo Sanz, Othalia Larue, Hedda R. Schmidtke, Sang Wan Lee, Keith McGreggor
P23	Non-ordinary Consciousness for Artificial Intelligence	Gabriel Axel Montes
P24	Promises and Limitations of Conscious Machines	Lalit M Patnaik, Jagadish Kallimani
P25	Reality Construction in Cognitive Agents Through Processes of Info-computation	Gordana Dodig-Crnkovic, Rickard von Haugwitz
P26	Reviewing Tests for Machine Consciousness	Aida Elamrani, Roman Yampolskiy
P27	The ant colony as a test for scientific theories of consciousness	Daniel A Friedman, Eirik Søvik
P28	The Computational Explanatory Gap	James A. Reggia, Derek Monner, Jared Sylvester
P29	The Cortical Conductor Theory Towards Addressing Consciousness in AI Models	Joscha Bach
P30	The Independent Core Observer Model Computational Theory of Consciousness and the Mathematical model for Subjective Experience	David Kelley
P31	The Information-Theoretic and Algorithmic Approach to Human, Animal, and Artificial Cognition	Nicolas Gauvrit, Héctor Zenil, Jesper Tegnér
P32	The multidimensional brain	Arturo Tozzi
P33	The rise of machine consciousness: Studying consciousness with computational models	James A. Reggia
P34	Towards an Affective Computational Model for Machine Consciousness	Rohitash Chandra
P35	Using Tononi Phi to Measure Consciousness of a Cognitive System While Reading and Conversing	Matthew Iklé, Ben Goertzel, Misgana Bayetta, George Sellman, Comfort Cover, Jennifer Allgeier, David R. Smith, Morris Sowards, Dylan Schuldberg, Man Hin Leung, A. Belayneh, Paul G. Smith, Andrew D. Hanson
P36	What are the computational correlates of consciousness?	James A. Reggia,Garrett Katz, Di-Wei Huang
P37	What is it Like to be a Group Agent?	Christian List
P38	Z-numbers Augmented Z-numbers for machine-subjectivity representation	Romi Banerjee, Sankar K. Pal
P39	The Blackwell Companion to Consciousness	Susan Schneider, Max Velmans
P40	Two theories of consciousness: Semantic pointer competition vs. information integration	Paul Thagard, Terrence Stewart

Id	Title	Authors
P41	Strategies for measuring machine consciousness	Raul Arrabales M., Agapito Ledezma E., Araceli Sanchis de Miguel
P42	Brain states matter. A reply to the unfolding argument	Johannes Kleiner
P43	Elements of qualitative cognition: An information topology perspective	Pierre Baudot
P44	Integrated information as a possible basis for plant consciousness	Paco Calvo, Frantisek Baluska, Anthony Trewavas
P45	A novel perturbation based compression complexity measure for networks	Mohit Virmani, Nithin Nagaraj
P46	Critical integration in neural and cognitive systems: Beyond power-law scaling as the hallmark of soft assembly	Miguel Aguilera,Ezequiel A. Di Paolo
P47	Efficient search for informational cores in complex systems: Application to brain networks	Jun Kitazono, Ryota Kanai,Masafumi Oizumi
P48	A macro agent and its actions	L. Albantakis, F. Massari, M. Beheler-Amass, G. Tononi
P49	An Integrated World Modeling Theory (IWMT) of Consciousness: Combining Integrated Information and Global Neuronal Workspace Theories With the Free Energy Principle and Active Inference Framework; Toward Solving the Hard Problem and Characterizing Agentic Causation	Adam Safron
P50	Artificial intelligence and its natural limits	K.D. Stephan, G. Klima
P51	A traditional scientific perspective on the integrated information theory of consciousness	J. Mallat
P52	Complexity as causal information integration	C. Langer, N. Ay
P53	Consciousness and Automated Reasoning	Ulrike BarthelmeÃ, Ulrich Furbach, Claudia Schon
P54	Calculating the boundaries of consciousness in general resonance theory	T. Hunt
P55	Finding continuity and discontinuity in fish schools via integrated information theory	T. Niizato, K. Sakamoto, Y. I. Mototake, H. Murakami, T.Tomaru, T. Hoshika, T.Fukushima
P56	How Morphological Computation Shapes Integrated Information in Embodied Agents	C. Langer, N. Ay
P57	The robot consciousness based on empirical knowledge	Yuanxiu Liao, Mingrui Yan, Suqin Tang
P58	Topological Model of Neural Information Networks	Matilde Marcolli
P59	The Emergence of Integrated Information, Complexity, and 'Consciousness' at Criticality	Nicholas J.M. Popiel, Sina Khajehabdollahi, Pubudita Abeyasinghe, Francesco Riganello, Emily Nichols, Adrian Owen, Andrea Soddu
P60	Consciousness: Just Another Technique?	U. BarthelmeÃ, U. Furbach
P61	How to catch an AI zombie: Testing for consciousness in machines	S. Schneider
P62	A mean field approach to model levels of consciousness from EEG recordings	Javarone, M. A.Gosseries, O. Marinazzo, D. Noirhomme, Q. Bonhomme, V. Laureys, S. Chennu, S
P63	AI-generated music: Creativity and autonomy	C. Moruzzi

Appendix 2

Dependent measure of the state	Dependent measure of the state	Features	Efficiency	Fuente
$\Phi_{3.0}$	$\Phi_{3.0}^{peak}$	Integrated information (II) according to IIT3.0	It is the measure according to IIT3.0 in order O(n53n) [95], for binary systems with n being the number of elements in the system	Oizumi [6]

Dependent measure of the state	Dependent measure of the state	Features	Efficiency	Fuente
CO $\Phi_{3.0}$	CO $\Phi_{3.0}^{peak}$	"Cut-one" Cut a connection when partitioning. Reduces the scope of the IPM search over all possible cuts in the system	It does not evaluate the partitioned CES for each of the 2n unidirectional bipartitions of the system, only the 2n bipartitions separating the arcs from a single node to the rest of the network or vice versa are evaluated. Since one wants to find the minimum value of Φ in all possible partitions, this approach gives an upper bound on the exact value of Φ of the system	Mayner et al. [95]
NN$\Phi_{3.0}$	NN$\Phi_{3.0}^{peak}$	"No new concepts" No new concepts after partitioning. It is possible that the CES (of the partitioned system) in most of the alternative mechanism partitioning schemes and distance measurements, of the partitioned system has reducible concepts in the non-partitioned system, thus, they would not be part of the non-partitioned CES. Thus, PyPhi calculates by default the partitioned CES from the beginning of the partitioned TPM. With this approach these new concepts are ignored. Only those mechanisms that already specify concepts in the non-partitioned CES are taken, and the entire CES calculation is not repeated for each partition of the system (which must reevaluate all possible candidate mechanisms for irreducibility)	In many types of systems, new concepts due to partitioning are rare. Therefore, approximations using the "no new concepts" option usually New concepts due to partitioning are rare in a wide variety of systems. Consequently, using "No new concepts" yields results that are usually accurate for approaches that use it. However, it should be clear that this approximation provides neither a theoretical upper bound nor a theoretical lower bound on the exact value of Φ of the system	Mayner et al. [95]
WS$\Phi_{3.0}$	WS$\Phi_{3.0}^{peak}$	System as a whole. All nodes in the system are included in the CM	Defining the complete system as the larger complex avoids testing each candidate subsystem	Nilsen et al. [64]
IC$\Phi_{3.0}$	IC$\Phi_{3.0}^{peak}$		Excluding these nodes is actually not an approximation but a shorter route since those nodes should be outside the larger complex	Nilsen et al. [64]
	Est.n$\Phi_{3.0}^{peak}$	Iterative cuts. The subsystem of the network where all nodes with non-recursive connectivity or an unreachable state are iteratively removed are included in the larger complex	It takes a few states instead of taking the maximum over all possible states. Its growth is exponential in computational time	Nilsen et al. [64]
$\Phi_{2.0}$	$\Phi_{.2.0}^{peak}$	Estimation of Φ3.0peak from n states (n = 1,2, 3,…,15)	Measure bounded by exponential growth in computational time	Balduzzi and Tononi [27]
$\Phi_{2.5}$	$\Phi_{.2.5}^{peak}$	Integrated information according to IIT2.0	Φ2.0 incorporating minimization over both cause-effect and not just cause. Measurement limited by exponential growth in computational time	Tegmark [96]
	D1	A combination between Φ2.0 and Φ3.0 features	It is inversely related to a degeneracy of the system of state transitions	Marshall et al. [97]
	D2	Achievable states. The number of states that were reachable is calculated by quantifying the repertoire of available states of the system	Indicates the degree of difference between system states	Marshall et al. [97]
	S	Cumulative variance of activity elements at each system node given the maximum entropy distribution of initial conditions	Heuristic measure for observed time series data. Used in EEG to distinguish states of consciousness	Schartner et al. [98]
	LZ	Sample coalition entropy. A measure of the entropy of the observed state distribution that indicates the average diversity of a system of visited states	Heuristic measure for observed time series data. Indicates the degree of order or patterns in sequences of observed states of a system. Used in EEG to distinguish states of consciousness	Schartner et al. [98]

Dependent measure of the state	Dependent measure of the state	Features	Efficiency	Fuente
	Φ^*	Functional complexity. Signal complexity measured by algorithmic compressibility through Lempel–Ziv compression	Employ an exhaustive search of the MIP with a bipartition scheme (2n-1–1)	Oizumi et al. [99]
	SI	Integrated decoder-based information on IIT2.0	Employ an exhaustive search of the IPM with a bipartition scheme (2n–1–1)	Barrett and Seth [108]
	MI	Integrated stochastic interaction. Based on IIT2.0	Employ an exhaustive IPM search with a bipartitioning scheme (2n–1–1)	Oizumi et al. [100]
	Geometrical integrated information	Mutual information. Based on IIT1.0	Enables methodological unification of many information theoretic measures such as transfer entropy, MI, SI and II. Quantifies the strength of multiple causal influences between elements, projecting the probability distribution of complex systems in restricted multidimensional varieties	Tozzi [20]

Appendix 3

Data of main authors are available in the link:
https://docs.google.com/spreadsheets/d/1A9c5QOsTL1siMXZyVin_8Va0u7WJOOeh/edit?usp=share_link&ouid=112201640395350008286&rtpof=true&sd=true

In addition, the main image is presented below with all the authors who published at least one related article.

Data availability statement The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no conflict of interest with this proposal.

References

1. Patnaik LM, Kallimani JS (2017) Promises and limitations of conscious machines. In: Menon S, Nagaraj N, Binoy VV (eds) Self, culture and consciousness: interdisciplinary convergences on knowing and being. Springer Singapore, Singapore
2. Crick F (1995) Astonishing hypothesis: the scientific search for the soul, new york, scribner reprint edition
3. Penrose R (1996) Shadows of the mind: a search for the missing science of consciousness United States of America. Oxford University Press, Oxford
4. Chalmers DJ (1995) Facing up to the problem of consciousness. J Conscious Stud 2:19
5. Tononi G (2004) An information integration theory of consciousness. BMC Neurosci 5:42
6. Oizumi M, Albantakis L, Tononi G (2014) From the phenomenology to the mechanisms of consciousness: integrated information theory 3.0. Plos Comput Biol 10(5):e1003588
7. Nilsson NJ (2009) The quest for artificial intelligence. Cambridge University Press
8. Holland O, Gamez D (2009) Artificial intelligence and consciousness. In: Banks WP (ed) Encyclopedia of consciousness. Academic Press
9. Stephan KD, Klima G (2021) Artificial intelligence and its natural limits. AI Soc 36:9–18
10. Nazri A, Ghani AAA, Hafez I, Ng KY (2018) A new theoretical framework for testing consciousness in a machine. In: International conference on soft computing and data mining. Springer, pp 330–339

11. Kriegeskorte N, Douglas PK (2018) Cognitive computational neuroscience. Nat Neurosci 21:1148–1160
12. Schwartz EL (1990) Computational neuroscience. MIT press
13. Marte H (2020) Neurociencia computacional: el futuro de la investigación [online]. Available: https://neuro-class.com/neurociencia-computacional-el-futuro-de-la-investigacion/ Accessed 2020
14. Makin JG, Moses DA, Chang EF (2020) Machine translation of cortical activity to text with an encoder–decoder framework. Nat Neurosci 23(4): 575–582
15. Wendin G (2019) Can biological quantum networks solve np-hard problems? Adv Quantum Technol 2(7–8):1800081
16. Russell S, Norvig P (2020) Artificial intelligence: a modern approach, Hoboken
17. Liao Y, Yan M, Tang S (2021, March). The robot consciousness based on empirical knowledge. In Journal of Physics: Conference Series (Vol. 1861, No. 1, p. 012103). IOP Publishing
18. Buczak AL, Guven E (2015) A survey of data mining and machine learning methods for cyber security intrusion detection. IEEE Commun Surv Tutor 18(2):1153–1176
19. Lecun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521:436–444
20. Tozzi A (2019) The multidimensional brain. Phys Life Rev 31:86–103
21. Gauvrit N, Zenil H, Tegnér J (2017) The information-theoretic and algorithmic approach to human, animal, and artificial cognition. Representation and reality in humans, other living organisms and intelligent machines 117–139
22. Kralik J, Lee JH, Rosenbloom PS, Jackson PC, Epstein SL, Romero OJ, Sanz R, Larue O, Schmidtke HR, Lee SW, Mcgreggor K (2018) Metacognition for a common model of cognition. Proc Comput Sci 145:730–739
23. Chandra R (2017) Towards an affective computational model for machine consciousness. In: Neural Information Processing: 24th International Conference, ICONIP 2017, Guangzhou, China, November 14–18, 2017, Proceedings, Part V 24 (pp. 897–907). Springer International Publishing
24. Block N (2009) Comparing the major theories of consciousness. In: The cognitive neurosciences, 4th edn. Massachusetts Institute of Technology, Cambridge.
25. Manzotti R (2019) A physicalist causally oriented foundation for a conscious machine based on the spread mind. In: AAAI spring symposium: towards conscious AI systems
26. Koch C, Tononi G (2008) Can machines be conscious? IEEE Spectr 45:55–59
27. Balduzzi D, Tononi G (2008) Integrated information in discrete dynamical systems: motivation and theoretical framework. Plos Comput Biol 4:e1000091
28. Tononi G, Koch C (2015) Consciousness: here, there and everywhere? Philos Trans R Soc Lond Ser B Biol Sci 370(1668):20140167
29. Tononi G, Boly M, Massimini M, Koch C (2016) Integrated information theory: from consciousness to its physical substrate. Nat Rev Neurosci 17:450–461
30. Paz F, Pow S (2016) A systematic mapping review of usability evaluation methods for software development process. Int J Softw Eng I Appl 10(1):165–178
31. Petersen K, Feldt R, Mujtaba S, Mattsson M (2008) Systematic mapping studies in software engineering. In: Proceedings of the 12th international conference on evaluation and assessment in software engineering, vol 17
32. Petersen K, Vakkalanka S, Kuzniarz L (2015) Guidelines for conducting systematic mapping studies in software engineering: an update. Inform Softw Technol 64:18
33. Kitchenham B, Brereton P (2013) A systematic review of systematic review process research in software engineering. Inf Softw Technol 55(12): 2049–2075
34. Franco-bedoya O, Ameller D, Costal D, Franch X (2017) Open source software ecosystems: a systematic mapping. Inform Softw Technol 91:160–185
35. Manikas, k. (2016) Revisiting software ecosystems research: a longitudinal literature study. J Syst Softw 117:84–103
36. Kitchenham B, Pearl Brereton O, Budgen D, Turner M, Bailey J, Linkman S (2009) Systematic literature reviews in software engineering – a systematic literature review. Inform Softw Technol 51:7–15
37. Miles MB, Huberman AM, Saldana J (2019) Qualitative data analysis: a methods sourcebook. Sage Publications
38. Kitchenham B, Brereton P (2013) A systematic review of systematic review process research in software engineering. Inf Softw Technol 55:2049–2075
39. McGreggor K (2020). An experience is a knowledge representation. Common Model of Cognition Bulletin 1(2): 420–424
40. Antonopoulos C, Fokas A, Bountis T (2015) Dynamical complexity in the c.elegans neural network. Eur Phys J Spec Top 225:1255–1269
41. Moreno RA., de Miguel AS, Espino AL (2011). Evaluation and development of consciousness in artificial cognitive systems. Ph. D. dissertation
42. Montes GA (2017) Non-ordinary consciousness for artificial intelligence. Springer International Publishing, pp 348–362
43. Friedman DA., Søvik E (2021) The ant colony as a test for scientific theories of consciousness. Synthese 198(2), 1457–1480
44. Reggia J, Monner D, Sylvester J (2014) The computational explanatory gap. J Conscious Stud 21(9–10):153–178
45. Banerjee R, Pal SK (2015) Z*-numbers: augmented z-numbers for machine-subjectivity representation. Inf Sci 323:143–217
46. Niizato T, Sakamoto K, Mototake YI, Murakami H, Tomaru T, Hoshika T, Fukushima T (2020) Finding continuity and discontinuity in fish schools via integrated information theory. PLoS ONE 15(2):e0229573
47. Marcolli M (2021) Topological model of neural information networks. Springer International Publishing, 623–633
48. Javarone MA, Gosseries O, Marinazzo D, Noirhomme Q, Bonhomme V, Laureys S, Chennu S (2020) A mean field approach to model levels of consciousness from eeg recordings. J Stat Mech Theory Exp 2020(8):083405
49. Reggia J (2013) The rise of machine consciousness: studying consciousness with computational models. Neural Netw Off J Int Neural Netw Soc 44:112–131
50. Iklé M, Goertzel B, Bayetta M, Sellman G, Cover C, Allgeier J, Smith R, Sowards M, Schuldberg D, Leung M, Belayneh A, Smith G, Hanson D (2019) Using tononi phi to measure consciousness of a cognitive system while reading and conversing [online]. https://ceur-ws.org/Vol-2287/paper20.pdf Accessed 2020
51. Popiel NJM, Khajehabdollahi S, Abeyasinghe PM, Riganello F, Nichols ES, Owen AM, Soddu A (2020) The emergence of integrated information, complexity, and 'consciousness' at criticality. Entropy 22:339
52. Farnsworth KD (2018) How oganisms gained causal independence and how it might be quantified. Biology 7:38
53. Kalita P, Langa JA, Soler-Toscano F (2019) Informational structures and informational fields as a prototype for the description of postulates of the integrated information theory. Entropy 21:493
54. Dodig-Crnkovic G, Von Haugwitz R (2017) Reality construction in cognitive agents through processes of info-computation. In: Dodig-Crnkovic G, Giovagnoli R (eds) Representation and

reality in humans, other living organisms and intelligent machines. Springer International Publishing, Cham
55. Reggia JA, Katz G, Huang DW (2016) What are the computational correlates of consciousness? Biol Inspir Cognit Arch 17:101–113
56. Arrabales R, Ledezma A, Sanchis A (2009) Strategies for measuring machine consciousness. Int J Mach Conscious 01:193–201
57. Aguilera M, Di Paolo EA (2021) Critical integration in neural and cognitive systems: beyond power-law scaling as the hallmark of soft assembly. Neurosci Biobehav Rev 123:230–237
58. Hunt T (2020) Calculating the boundaries of consciousness in general resonance theory. J Conscious Stud 27:55–80
59. Langer C, Ay N (2021) How morphological computation shapes integrated information in embodied agents. Frontiers in Psychol. 5385
60. Schneider S (2020) How to catch an ai zombie: testing for consciousness in machines. Ethics of artificial intelligence
61. Koch C, Tononi G (2011) A test for consciousness. Sci Am 304:44–47
62. Arrabales R, Ledezma A, Sanchis A (2010) ConsScale: A pragmatic scale for measuring the level of consciousness in artificial agents. J Consciousness Stud 17(3-4), 131–164
63. Pizzi R, Musumeci M (2017) Coding Mental States from EEG Signals and evaluating their Integreted Information Content: a Computational Intelligence Approach. Int J Circ Syst Signals Process 11(4464):464–470
64. Nilsen AS, Juel B, Marshall W (2019) Evaluating approximations and heuristic measures of integrated information. Entropy 21:525
65. Elamrani A, Yampolskiy RV (2019) Reviewing tests for machine consciousness. J Conscious Stud 26:35–64
66. Kanade VA (2020) A hybrid evolutionary algorithm for evolving a conscious machine. In Intelligent Systems Design and Applications: 18th International Conference on Intelligent Systems Design and Applications (ISDA 2018) held in Vellore, India, December 6-8, 2018, Volume 2 (pp. 1045–1054). Springer International Publishing
67. Kitazono J, Kanai R, Oizumi M (2020) Efficient search for informational cores in complex systems: application to brain networks. Neural Netw 132:232–244
68. Virmani M, Nagaraj N (2019) A novel perturbation based compression complexity measure for networks. Heliyon 5:e01181
69. Barghout L (2018) Hypernym and spatial-taxon hierarchy. A cognitive informatics and fuzzy logic approach to combining linguistic and image taxonomies. In: 2018 IEEE 17th international conference on cognitive informatics and cognitive computing (ICCI* CC), pp 575–582. IEEE
70. Van Hateren JH (2019) A theory of consciousness: computation, algorithm, and neurobiological realization. Biol Cybern 113:357–372
71. Shanahan M (2015) Ascribing consciousness to artificial intelligence. arXiv preprint arXiv:1504.05696
72. Findlay G, Marshall W, Albantakis L, Mayner W, Koch C, Tononi G (2019). Dissociating intelligence from consciousness in artificial systems–Implications of Integrated Information Theory. In Proceedings of the 2019 Towards Conscious AI Systems Symposium, AAAI SSS19
73. Bedia M, Castillo L (2016) Hacia una teoría de la mente corporizada: la influencia de los mecanismos sensomotores en el desarrollo de la cognición. Ánfora 17:101–124
74. Montemayor C, de Barros JA, de Assis lPG, (2019) Implementation, formalization, and representation: challenges for integrated information theory. J Conscious Stud 26:107–132
75. Bach J (2019) The cortical conductor theory: Towards addressing consciousness in ai models. In Biologically Inspired Cognitive Architectures 2018: Proceedings of the Ninth Annual Meeting of the BICA Society (pp. 16–26). Springer International Publishing
76. Kelley D (2018) The independent core observer model computational theory of consciousness and the mathematical model for subjective experience. In: 2018 4th annual international conference on network and information systems for computers (ICNISC), 19–21 April 2018, pp 396–400
77. List C (2018) What is it like to be a group agent? Noûs 52:295–319
78. Schneider S, Velmans M (2017) The Blackwell Companion to Consciousness. Blackwell Publishing, 175–179
79. Thagard P, Stewart TC (2014) Two theories of consciousness: semantic pointer competition versus Information integration. Conscious Cognit 30:73–90
80. Calvo P, Baluška F, Trewavas A (2021) Integrated information as a possible basis for plant consciousness. Biochem Biophys Res Commun 564:158–165
81. Kleiner J (2020) Brain states matter: a reply to the unfolding argument. Conscious Cognit 85:102981
82. Baudot P (2019) Elements of qualitative cognition: an information topology perspective. Phys Life Rev 31:263–275
83. Safron A (2020) An integrated world modeling theory (iwmt) of consciousness: combining integrated information and global neuronal workspace theories with the free energy principle and active inference framework; toward solving the hard problem and characterizing agentic causation. Front Artif Intell 3:30
84. Mallatt J (2021) A traditional scientific perspective on the integrated information theory of consciousness. Entropy 23(6):650
85. Barthelmeß U, Furbach U, Schon C (2020) Consciousness and Automated Reasoning. arXiv preprint https://doi.org/10.48550/arXiv.2001.09442
86. Moruzzi C (2019) Ai-generated music: creativity and autonomy. pp 25–32
87. Barrett A, Seth AK (2011) Practical measures of integrated information for time-series data. Plos Comput Biol 7:e1001052
88. Hoel EP, Albantakis L, Marshall W, Tononi G (2016) Can the macro beat the micro? Integrated information across spatiotemporal scales. Neurosci of Conscious 2016(1):1–13
89. Marshall W, Albantakis L, Tononi G (2018) Black-boxing and cause-effect power. PLoS Comput Biol 14:e1006114
90. Albantakis L, Massari F, Beheler-Amass M, Tononi G (2021) A macro agent and its actions. In Top-Down Causation and Emergence (pp. 135–155). Cham: Springer International Publishing
91. Gamez D (2010) Information integration based predictions about the conscious states of a spiking neural network. Conscious Cogn 19:294–310
92. Sylvester JC, Reggia JA, Weems SA, Bunting MF (2013) Controlling working memory with learned instructions. Neural Netw 41:23–38
93. Aleksander I, Gamez D (2009) Iconic training and effective information: evaluating meaning in discrete neural networks. AAAI fall symposium - technical report
94. Karl, F. (2012). A free energy principle for biological systems. Entropy 14(11): 2100–2121
95. Mayner WGP, Marshall W, Albantakis L, Findlay G, Marchman R, Tononi G (2018) Pyphi: a toolbox for integrated information theory. Plos Comput Biol 14:e1006343
96. Tegmark, m. (2016) Improved measures of integrated information. PLoS Comput Biol 12:e1005123
97. Marshall W, Gomez-ramirez J, Tononi G (2016) Integrated information and state differentiation. Front Psychol 7:926

98. Schartner M, Seth A, Noirhomme Q, Boly M, Bruno M-A, Laureys S, Barrett A (2015) Complexity of multi-dimensional spontaneous EEG decreases during propofol induced general anaesthesia. PLoS ONE 10:e0133532
99. Oizumi M, Amari S, Yanagawa T, Fujii N, Tsuchiya N (2016) Measuring integrated information from the decoding perspective. PLoS Comput Biol 12:e1004654–e1004654
100. Oizumi, m., tsuchiya, n. & amari, s. I. (2016) Unified framework for information integration based on information geometry. Proc Natl Acad Sci 113:14817–14822
101. Kitazono J, Oizumi M (2018) Practical phi toolbox
102. Krohn, S., Ostwald, D. (2017). Computing integrated information. Neuroscience of consciousness, 3(1): 1-16.
103. Aguilera M, di Paolo AE (2019) Integrated information in the thermodynamic limit. Neural netw 114:136–146
104. Tajima S, Kanai R (2017) Integrated information and dimensionality in continuous attractor dynamics. Neurosci Conscious 2017(1):nix011
105. Bach J (2019) The cortical conductor theory: towards addressing consciousness in AI models. In: Biologically inspired cognitive architectures 2018
106. Aguilera, M. (2019). Scaling behaviour and critical phase transitions in integrated information theory. Entropy 21(12):1198. https://doi.org/10.3390/e21121198
107. Tononi G, Sporns O (2003) Measuring information integration. BMC Neurosci 4:31
108. Barrett A, Seth AK (2011) Practical measures of integrated information for time-series data. PLoS Comput Biol 7:e1001052

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.