WIKIPEDIA

Intelligent agent

In artificial intelligence, an **intelligent agent** (**IA**) refers to an autonomous entity which acts, directing its activity towards achieving goals (i.e. it is an agent), upon an environment using observation through sensors and consequent actuators (i.e. it is intelligent).[1] Intelligent agents may also learn or use knowledge to achieve their goals. They may be very simple or very complex. A reflex machine, such as a thermostat, is considered an example of an intelligent agent.[2]

Intelligent agents are often described schematically as an abstract functional system similar to a computer program. Researchers such as Russell & Norvig (2003) consider goal-directed behavior to be the essence of intelligence; a normative agent can be labeled with a term borrowed from economics, "rational agent". In this rational-action paradigm, an IA possesses an internal "model" of its environment. This model encapsulates all the agent's beliefs about the world. The agent also has an "objective function" that encapsulates all the IA's goals. Such an agent is designed to create and execute whatever plan will, upon completion, maximize the expected value of the objective function.[3] A reinforcement learning agent can have a "reward function" that allows the programmers to shape the IA's desired behavior,[4] and an evolutionary algorithm's

Simple reflex agent

behavior is shaped by a "fitness function".[5] Abstract descriptions of intelligent agents are sometimes called **abstract intelligent agents** (**AIA**) to distinguish them from their real world implementations as computer systems, biological systems, or organizations. Some **autonomous intelligent agents** are designed to function in the absence of human intervention. As intelligent agents become more popular, there are increasing legal risks involved.[6]:815

Intelligent agents in artificial intelligence are closely related to agents in economics, and versions of the intelligent agent paradigm are studied in cognitive science, ethics, the philosophy of practical reason, as well as in many interdisciplinary socio-cognitive modeling and computer social simulations.

Intelligent agents are also closely related to software agents (an autonomous computer program that carries out tasks on behalf of users). In computer science, an *intelligent agent* is a software agent that has some intelligence, for example, autonomous programs used for operator assistance or data mining (sometimes referred to as *bots*) are also called "intelligent agents".
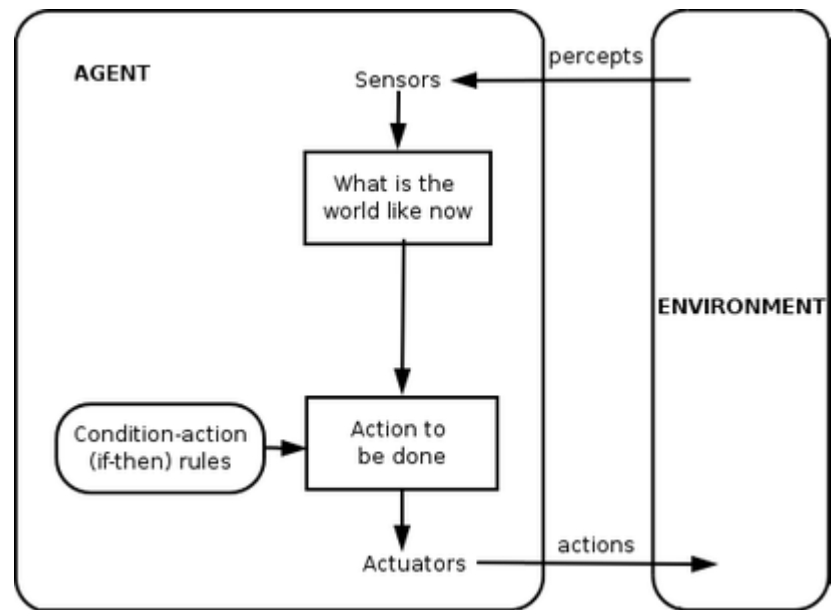
Contents

Definitions and characteristics

According to Nikola Kasabov (1998),[7] IA systems should exhibit the following characteristics:

- Accommodate new problem solving rules incrementally
- Adapt online and in real time
- Are able to analyze themselves in terms of behavior, error and success.
- Learn and improve through interaction with the environment (embodiment)
- Learn quickly from large amounts of data
- Have memory-based exemplar storage and retrieval capacities
- Have parameters to represent short and long term memory, age, forgetting, etc.

Padgham & Winikoff (2005) agree that an intelligent agent is situated in an environment and responds (in a timely, though not necessarily real-time) manner to environment changes. However, intelligent agents must also proactively pursue goals in a flexible and robust way.[a] Optional desiderata include that the agent be rational, and that the agent be capable of belief-desire-intention analysis.[8] Some 20th-century definitions characterize an agent as a program that aids a user or that acts on behalf of a user.[9] The influential *AIMA* (2009) defines an agent as "anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators", and characterizes intelligence as the ability to succeed at acting in accordance with certain ideal standards for rationality.[10][11]

"Intelligent agent" is also often used as a vague marketing term, sometimes synonymous with "virtual personal assistant".[12]

Objective function

Some agents can be assigned an explicit "goal function"; an agent is considered more intelligent if it consistently takes actions that successfully maximize its programmed goal function. The "goal function" encapsulates all of the goals the agent is driven to act on; in the case of rational agents, the function also encapsulates the acceptable trade-offs between accomplishing conflicting goals. (Terminology varies; for example, some agents seek to maximize or minimize a "utility function", "objective function", or "loss function".)[10][11] The theoretical and uncomputable AIXI design is a maximally intelligent agent in this paradigm;[13] however, in the real world, the IA is constrained by finite time and hardware resources, and scientists compete to produce algorithms that can achieve progressively higher scores on benchmark tests with real-world hardware.[14]

Systems that are not traditionally considered agents, such as knowledge-representation systems, are sometimes subsumed into the paradigm by framing them as agents that have a goal of (for example) answering questions as accurately as possible; the concept of an "action" is here extended to encompass the "act" of giving an answer to a question. As an additional extension, mimicry-driven systems can be framed as agents who are optimizing a "goal function" based on how closely the IA succeeds in mimicking the desired behavior.[10][11] In the generative adversarial networks of the 2010s, an "encoder"/"generator" component attempts to mimic and improvise human text composition. The generator is attempting to maximize a function encapsulating how well it can fool an antagonistic "predictor"/"discriminator" component.[15]

While GOFAI systems often accept an explicit goal function, the paradigm can also be applied to neural networks and to evolutionary computing. Reinforcement learning can generate intelligent agents that appear to act in ways intended to maximize a "reward function".[16] Sometimes, rather than setting the reward function to be directly equal to the desired benchmark evaluation function, machine learning programmers will use reward shaping to initially give the machine rewards for incremental progress in learning.[17] Yann LeCun stated in 2018 that "Most of the learning algorithms that people have come up with essentially consist of minimizing some objective function."[18] AlphaZero chess had a simple objective function; each win counted as +1 point, and each loss counted as -1 point. An objective function for a self-driving car would have to be more complicated.[19] Evolutionary computing can evolve intelligent agents that appear to act in ways intended to maximize a "fitness function" that influences how many descendants each agent is allowed to leave.[20]

Structure of agents

A simple agent program can be defined mathematically as a function f (called the "agent function")[21] which maps every possible percepts sequence to a possible action the agent can perform or to a coefficient, feedback element, function or constant that affects eventual actions:

$$f : P^* \to A$$

Agent function is an abstract concept as it could incorporate various principles of decision making like calculation of utility of individual options, deduction over logic rules, fuzzy logic, etc.[22]

The program agent, instead, maps every possible percept to an action.[23]

We use the term percept to refer to the agent's perceptional inputs at any given instant. In the following figures, an agent is anything that can be viewed as perceiving its environment through sensors and acting upon that environment through actuators.

Architectures

Weiss (2013) defines four classes of agents:

- Logic-based agents – in which the decision about what action to perform is made via logical deduction;
- Reactive agents – in which decision making is implemented in some form of direct mapping from situation to action;
- Belief-desire-intention agents – in which decision making depends upon the manipulation of data structures representing the beliefs, desires, and intentions of the agent; and finally,
- Layered architectures – in which decision making is realized via various software layers, each of which is more or less explicitly reasoning about the environment at different levels of abstraction.

Generally, an agent can be constructed by separating the body into the sensors and actuators, and so that it operates with a complex perception system that takes the description of the world as input for a controller and outputs commands to the actuator. However, a hierarchy of controller layers is often necessary to balance the immediate reaction desired for low-level tasks and the slow reasoning about complex, high-level goals.[24]

Classes

Russell & Norvig (2003) group agents into five classes based on their degree of perceived intelligence and capability:[25]

1. simple reflex agents
2. model-based reflex agents
3. goal-based agents
4. utility-based agents
5. learning agents
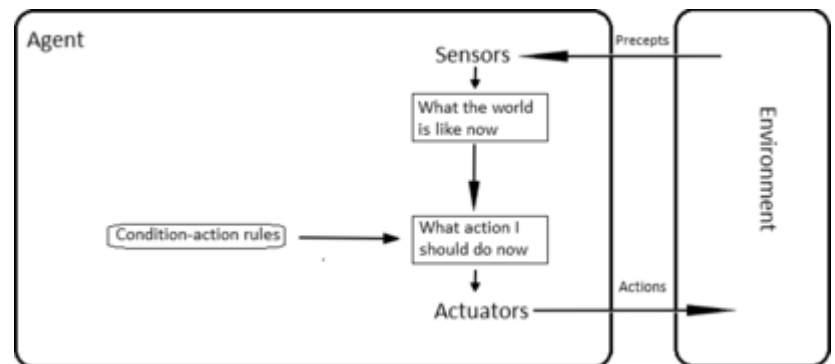
Simple reflex agent

Simple reflex agents

Simple reflex agents act only on the basis of the current percept, ignoring the rest of the percept history. The agent function is based on the *condition-action rule*: "if condition, then action".

This agent function only succeeds when the environment is fully observable. Some reflex agents can also contain information on their current state which allows them to disregard conditions whose actuators are already triggered.
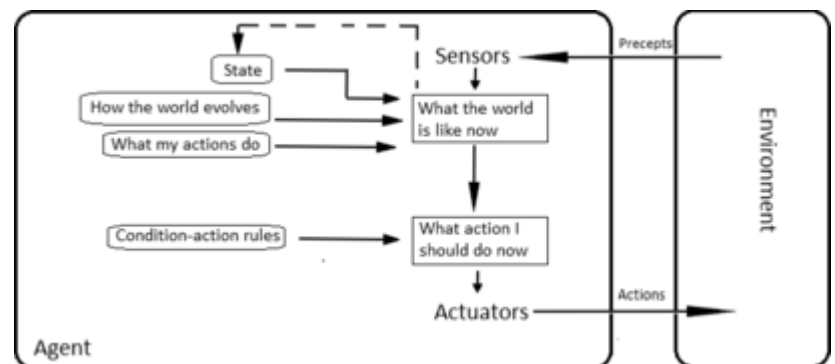
Model-based reflex agent

Infinite loops are often unavoidable for simple reflex agents operating in partially observable environments. Note: If the agent can randomize its actions, it may be possible to escape from infinite loops.

Model-based reflex agents

A model-based agent can handle partially observable environments. Its current state is stored inside the agent maintaining some kind of structure that describes the part of the world which cannot be seen. This knowledge about "how the world works" is called a model of the world, hence the name "model-based agent".

A model-based reflex agent should maintain some sort of internal model that depends on the percept history and thereby reflects at least some of the unobserved aspects of the current state. Percept history and impact of action on the environment can be determined by using the internal model. It then chooses an action in the same way as reflex agent.

An agent may also use models to describe and predict the behaviors of other agents in the environment.[26]

Model-based, goal-based agent
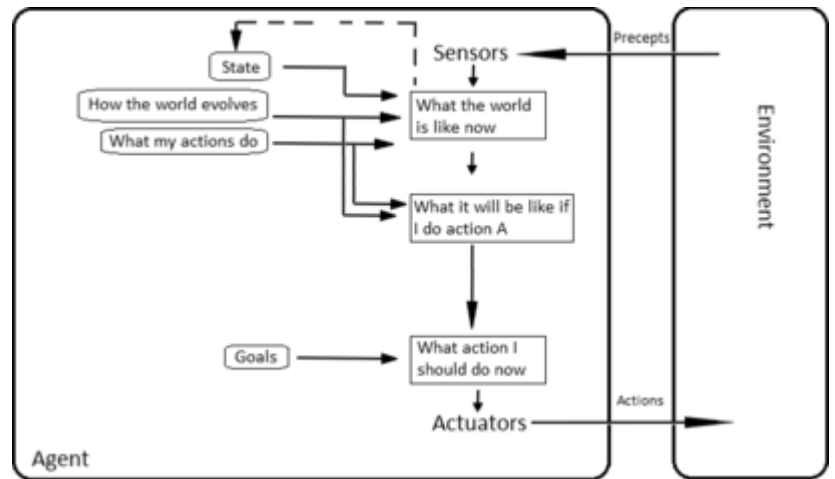
Goal-based agents

Goal-based agents further expand on the capabilities of the model-based agents, by using "goal" information. Goal information describes situations that are desirable. This provides the agent a way to choose among multiple possibilities, selecting the one which reaches a goal state. Search and planning are the subfields of artificial intelligence devoted to finding action sequences that achieve the agent's goals.

Model-based, utility-based agent
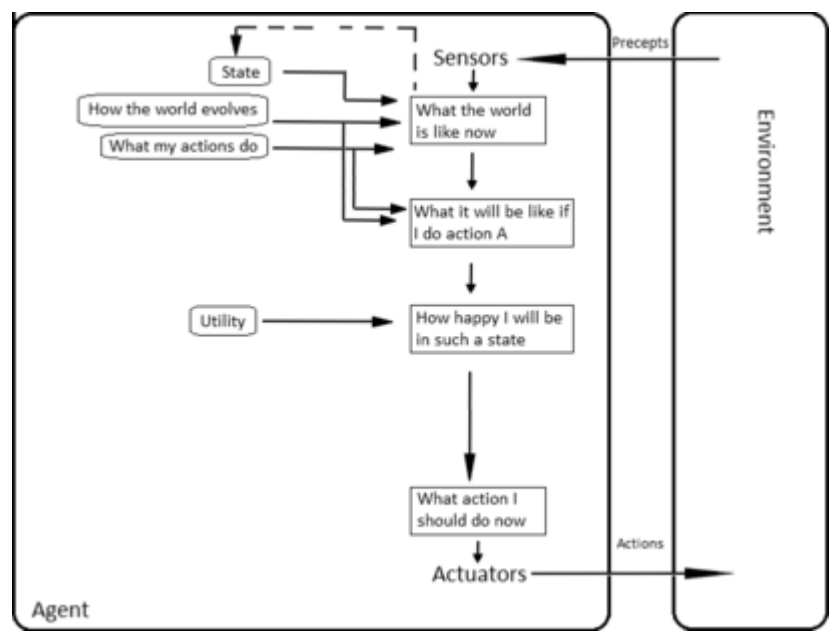
Utility-based agents

Goal-based agents only distinguish between goal states and non-goal states. It is also possible to define a measure of how desirable a particular state is. This measure can be obtained through the use of a *utility function* which maps a state to a measure of the utility of the state. A more general performance measure should allow a comparison of different world states according to how well they satisfied the agent's goals. The term utility can be used to describe how "happy" the agent is.

A rational utility-based agent chooses the action that maximizes the expected utility of the action outcomes - that is, what the agent expects to derive, on average, given the probabilities and utilities of each outcome. A utility-based agent has to model and keep track of its environment, tasks that have involved a great deal of research on perception, representation, reasoning, and learning.

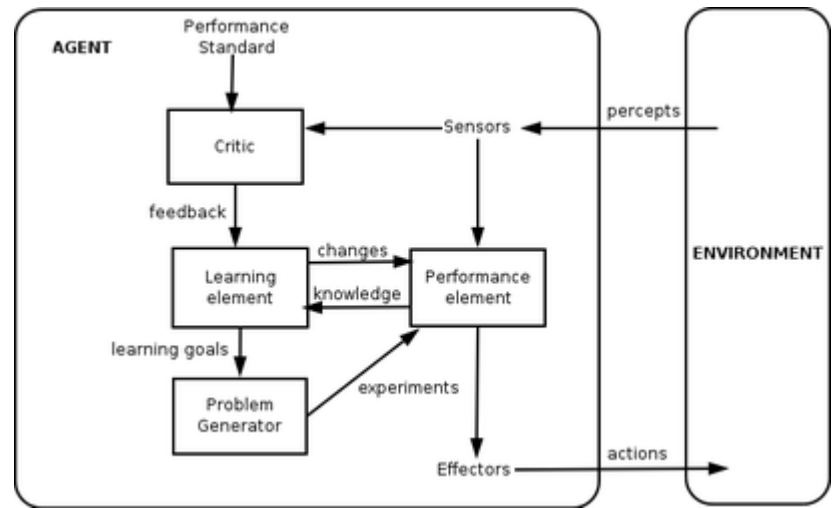## Learning agents

Learning has the advantage that it allows the agents to initially operate in unknown environments and to become more competent than its initial knowledge alone might allow. The most important distinction is between the "learning element", which is responsible for making improvements, and the "performance element", which is responsible for selecting external actions.

The learning element uses feedback from the "critic" on how the agent is doing and determines how the performance element, or "actor", should be modified to do better in the future. The performance element is what we have previously considered to be the entire agent: it takes in percepts and decides on actions.

A general learning agent

The last component of the learning agent is the "problem generator". It is responsible for suggesting actions that will lead to new and informative experiences.

Hierarchies of agents

To actively perform their functions, Intelligent Agents today are normally gathered in a hierarchical structure containing many "sub-agents". Intelligent sub-agents process and perform lower level functions. Taken together, the intelligent agent and sub-agents create a complete system that can accomplish difficult tasks or goals with behaviors and responses that display a form of intelligence.

Applications

Intelligent agents are applied as automated online assistants, where they function to perceive the needs of customers in order to perform individualized customer service. Such an agent may basically consist of a dialog system, an avatar, as well an expert system to provide specific expertise to the user.[27] They can also be used to optimize coordination of human groups online.[28]

An example of an automated online assistant providing automated customer service on a webpage.

See also

- Software agent
- Cognitive architectures
- Cognitive radio – a practical field for implementation
- Cybernetics, Computer science
- Data mining agent
- Embodied agent
- Federated search – the ability for agents to search heterogeneous data sources using a single vocabulary
- Fuzzy agents – IA implemented with adaptive fuzzy logic

- GOAL agent programming language
- Intelligence
- Intelligent system
- JACK Intelligent Agents
- Multi-agent system and multiple-agent system – multiple interactive agents
- PEAS classification of an agent's environment
- Reinforcement learning
- Semantic Web – making data on the Web available for automated processing by agents
- Simulated reality
- Social simulation
- Era of intelligent agents

Notes

a. The Padgham & Winikoff definition explicitly covers only social agents that interact with other agents.

Inline references

1. Anderson, Michael; Anderson, Susan Leigh (2007-12-15). "Machine Ethics: Creating an Ethical Intelligent Agent" (https://ojs.aaai.org/index.php/aimagazine/article/view/2065). *AI Magazine*. **28** (4): 15–15. doi:10.1609/aimag.v28i4.2065 (https://doi.org/10.1609%2Faimag.v28i4.2065). ISSN 2371-9621 (https://www.worldcat.org/issn/2371-9621).
2. According to the definition given by Russell & Norvig (2003, chpt. 2)
3. Bringsjord, Selmer and Govindarajulu, Naveen Sundar, "Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), URL = https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/.
4. Wolchover, Natalie (30 January 2020). "Artificial Intelligence Will Do What We Ask. That's a Problem" (https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/). *Quanta Magazine*. Retrieved 21 June 2020.
5. Bull, Larry. "On model-based evolutionary computation." Soft Computing 3, no. 2 (1999): 76-82.
6. Van Loo, Rory (2019-03-01). "Digital Market Perfection" (https://scholarship.law.bu.edu/faculty_scholarship/550). *Michigan Law Review*. **117** (5): 815.
7. Kasabov 1998
8. Lin Padgham and Michael Winikoff. Developing intelligent agent systems: A practical guide. Vol. 13. John Wiley & Sons, 2005.
9. Burgin, Mark, and Gordana Dodig-Crnkovic. "A systematic approach to artificial agents." arXiv preprint arXiv:0902.3513 (2009).
10. Russell & Norvig (2003)
11. Bringsjord, Selmer and Govindarajulu, Naveen Sundar, "Artificial Intelligence", The Stanford Encyclopedia of Philosophy (Summer 2020 Edition), Edward N. Zalta (ed.), forthcoming URL = <https://plato.stanford.edu/archives/sum2020/entries/artificial-intelligence/>.
12. Fingar, Peter (2018). "Competing For The Future With Intelligent Agents... And A Confession" (https://www.forbes.com/sites/cognitiveworld/2018/11/11/competing-for-the-future-with-intelligent-agents-and-a-confession/). *Forbes Sites*. Retrieved 18 June 2020.

13. Adams, Sam; Arel, Itmar; Bach, Joscha; Coop, Robert; Furlan, Rod; Goertzel, Ben; Hall, J. Storrs; Samsonovich, Alexei; Scheutz, Matthias; Schlesinger, Matthew; Shapiro, Stuart C.; Sowa, John (15 March 2012). "Mapping the Landscape of Human-Level Artificial General Intelligence" (https://doi.org/10.1609%2Faimag.v33i1.2322). *AI Magazine*. **33** (1): 25. doi:10.1609/aimag.v33i1.2322 (https://doi.org/10.1609%2Faimag.v33i1.2322).

14. Hutson, Matthew (27 May 2020). "Eye-catching advances in some AI fields are not real" (https://www.sciencemag.org/news/2020/05/eye-catching-advances-some-ai-fields-are-not-real). *Science | AAAS*. Retrieved 18 June 2020.

15. "Generative adversarial networks: What GANs are and how they've evolved" (https://venturebeat.com/2019/12/26/gan-generative-adversarial-network-explainer-ai-machine-learning/). *VentureBeat*. 26 December 2019. Retrieved 18 June 2020.

16. Wolchover, Natalie (January 2020). "Artificial Intelligence Will Do What We Ask. That's a Problem" (https://www.quantamagazine.org/artificial-intelligence-will-do-what-we-ask-thats-a-problem-20200130/). *Quanta Magazine*. Retrieved 18 June 2020.

17. Andrew Y. Ng, Daishi Harada, and Stuart Russell. "Policy invariance under reward transformations: Theory and application to reward shaping." In ICML, vol. 99, pp. 278-287. 1999.

18. Martin Ford. Architects of Intelligence: The truth about AI from the people building it. Packt Publishing Ltd, 2018.

19. "Why AlphaZero's Artificial Intelligence Has Trouble With the Real World" (https://www.quantamagazine.org/why-alphazeros-artificial-intelligence-has-trouble-with-the-real-world-20180221/). *Quanta Magazine*. 2018. Retrieved 18 June 2020.

20. Bull, Larry. "On model-based evolutionary computation." Soft Computing 3, no. 2 (1999): 76-82.

21. Russell & Norvig 2003, p. 33

22. Salamon, Tomas (2011). *Design of Agent-Based Models* (http://www.designofagentbasedmodels.info/). Repin: Bruckner Publishing. pp. 42–59. ISBN 978-80-904661-1-1.

23. Nilsson, Nils J. (April 1996). "Artificial intelligence: A modern approach" (https://dx.doi.org/10.1016/0004-3702(96)00007-0). *Artificial Intelligence*. **82** (1–2): 369–380. doi:10.1016/0004-3702(96)00007-0 (https://doi.org/10.1016%2F0004-3702%2896%2900007-0). ISSN 0004-3702 (https://www.worldcat.org/issn/0004-3702).

24. Poole, David; Mackworth, Alan. "1.3 Agents Situated in Environments▸ Chapter 2 Agent Architectures and Hierarchical Control▸ Artificial Intelligence: Foundations of Computational Agents, 2nd Edition" (https://artint.info/2e/html/ArtInt2e.Ch2.S3.html). *artint.info*. Retrieved 28 November 2018.

25. Russell & Norvig 2003, pp. 46–54

26. Stefano Albrecht and Peter Stone (2018). Autonomous Agents Modelling Other Agents: A Comprehensive Survey and Open Problems. Artificial Intelligence, Vol. 258, pp. 66-95. https://doi.org/10.1016/j.artint.2018.01.002

27. *Providing Language Instructor with Artificial Intelligence Assistant*. By Krzysztof Pietroszek. International Journal of Emerging Technologies in Learning (iJET), Vol 2, No 4 (2007) [1] (http://online-journals.org/i-jet/article/viewArticle/142) "Archived copy" (https://web.archive.org/web/20120307161853/http://www.doaj.org/doaj?func=abstract&id=239747). Archived from the original (http://www.doaj.org/doaj?func=abstract&id=239747) on 2012-03-07. Retrieved 2012-01-29.

28. Shirado, Hirokazu; Christakis, Nicholas A (2017). "Locally noisy autonomous agents improve global human coordination in network experiments" (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5912653). *Nature*. **545** (7654): 370–374. Bibcode:2017Natur.545..370S (https://ui.adsabs.harvard.edu/abs/2017Natur.545..370S). doi:10.1038/nature22332 (https://doi.org/10.1038%2Fnature22332). PMC 5912653 (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5912653). PMID 28516927 (https://pubmed.ncbi.nlm.nih.gov/28516927).

Other references

- Russell, Stuart J.; Norvig, Peter (2003). *Artificial Intelligence: A Modern Approach* (http://aima.cs.berkeley.edu/) (2nd ed.). Upper Saddle River, New Jersey: Prentice Hall. Chapter 2. ISBN 0-13-790395-2.
- Franklin, Stan; Graesser, Art (1996). "Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents" (https://web.archive.org/web/20150529072536/http://www.inf.ufrgs.br/~alvares/CMP124SMA/IsItAnAgentOrJustAProgram.pdf) (PDF). *Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages*. Springer-Verlag. Archived from the original (http://www.inf.ufrgs.br/~alvares/CMP124SMA/IsItAnAgentOrJustAProgram.pdf) (PDF) on 29 May 2015.
- Kasabov, N. (1998). "Introduction: Hybrid intelligent adaptive systems". *International Journal of Intelligent Systems*. **13** (6): 453–454. doi:10.1002/(SICI)1098-111X(199806)13:6<453::AID-INT1>3.0.CO;2-K (https://doi.org/10.1002%2F%28SICI%291098-111X%28199806%2913%3A6%3C453%3A%3AAID-INT1%3E3.0.CO%3B2-K).
- Weiss, G. (2013). *Multiagent systems* (2nd ed.). Cambridge, MA: MIT Press. ISBN 978-0-262-01889-0.

External links

- Coneural (http://www.coneural.org/reports/Coneural-03-01.pdf)