Towards Cognitive Bots:
Architectural Research Challenges

Habtom Kahsay Gidey[1], Peter Hillmann[1], Andreas Karcher[1], and Alois Knoll[2]

[1] Universität der Bundeswehr München, Germany
{habtom.gidey, peter.hillmann, andreas.karcher}@unibw.de
[2] Technische Universität München, München, Germany
{knoll}@in.tum.de

Abstract. Software bots operating in multiple virtual digital platforms must understand the platforms' affordances and behave like human users. Moreover, bots in such platforms could cooperate with humans or other software agents for work or to learn specific behavior patterns. Platform affordances or features change from one application platform to another or through a life cycle, requiring such bots to be adaptable. However, present-day bots, particularly chatbots, other than language processing and prediction, are far from reaching a human user's behavior level within complex business information systems. They lack the cognitive capabilities to sense and act in such virtual environments, rendering their development a challenge to artificial general intelligence research. In this study, we problematize and investigate assumptions in conceptualizing software bot architecture and direct attention to significant architectural research challenges in developing cognitive bots endowed with complex behavior, aka cognitive architectures, operating on information systems. As an outlook, we propose alternate architectural assumptions to consider in future bot design and bot development frameworks.

Keywords: cognitive bot · cognitive architecture · problematization.

1 Introduction

Bots are software agents that operate in digital virtual environments [33, 10]. An example scenario would be a "*user-like*" bot that could access web platforms as a user and behave like a human user. Ideally, such a bot could autonomously sense and understand the platforms' affordances. Affordances in digital spaces are, for example, interaction possibilities and functionalities on the web, in software services, or on web application platforms [42, 3, 35]. The bot would recognize and understand the differences and variability between different environments. If the platform or service has extensions to devices or bodies, as in the Web of Things (WoT), it would also have control or possibilities to interact with an outer web or service application world. Furthermore, a bot can be independent of a specific platform. A user-like social bot, for instance, would be able to recognize and understand social networks and act to influence or engage in belief sharing

on any social platform. It would also adjust with the changes and uncertainty of the affordances in a specific digital environment, such as when hypermedia interactivity features and functionalities change. Such a bot could also learn and develop to drive its goals and intentions from these digital microenvironments and take goal-directed targeted action to achieve them [17]. Such bots could also communicate and cooperate with other user agents, humans or bots, to collaborate and socialize for collective understanding and behavior.

The example scenarios described above convey the problems of perception and action in bots, similar to how a human user would perceive and act in digital spaces. To date, bots are incapable of the essential cognitive skills required to engage in such activity since this would entail complex visual recognition, language understanding, and the employment of advanced cognitive models. Most bots are instead conversational interfaces or question-and-answer knowledge agents, and others only perform automated repetitive tasks based on pre-given scripts, lacking autonomy and other advanced cognitive skills [14, 34]. The problems are, therefore, complex and challenging [41, 58]. Solutions must address different areas, such as transduction, autonomous action, and reasoning cases, to realize advanced generalizable intelligent behavior [18].

Problems spanning diverse domains require architectural solutions. Accordingly, these challenges also require researchers to address such systems' structural and dynamical elements from an architectural perspective [19, 45, 36]. For this reason, this paper aims to outline the architectural research agendas to address the problems toward conceptualizing and developing a cognitive bot with generalizable intelligence.

The paper is divided into sections discussing each of the research challenges. In Sect. 2, we discuss the challenges related to efforts and possible directions in enabling bots to sense and understand web platforms. Next, Sect. 3 describes the challenges related to realizing advanced cognitive models in software bots. Sect. 4 and 5, discuss the research issues in bot communication and cooperation, respectively. The remaining two sections provide general discussions on bot ethics and trust and conclude the research agenda.

## 2   The Transduction Problem

Web platforms can be seen as microenvironments or as distinct nature of digital microcosms [12]. They offer a microhabitat for their users' diverse digital experiences. These experiences mainly transpire from the elements of interaction and action, the hypermedia within web environments [12, 40]. Hypermedia connects and extends the experience, linking to further dimensions of the web-worlds, which means more pages and interactive elements from the user's perspective. Analogous to the biological concept of affordances from environmental pyschology [15], the interaction elements are considered affordances in the digital space [3, 42, 35]. Similarly, signifiers can also accompany affordances. Signifiers reveal or indicate apparent possibilities for actions associated with affordances [35, 52]. An example on the web would be a button affording a click

action and a text signifier hinting "Click to submit". A human user understands the web environment, its content or affordances, and navigates reasonably easily. However, enabling software bots to understand this digital environment and its affordances the way human users do is a challenging task. It is a complex problem of translating and mapping perception to action, i.e., the transduction problem [57, 2].

Today, there are different approaches to these challenges. The first category of approaches depends on providing knowledge about the environment for different levels of observability using APIs or knowledge descriptions. With API-based approaches, bots are developed for a specific platform, constantly putting developers in the loop. Bots do not have the general perceptual capability to understand and navigate with autonomous variability. Other architectures in this category, originating from the Web of Things (WoT), attempt to address the challenge using knowledge models and standards that could enable agents to perceive the web by exposing hypermedia affordances and signifiers [3, 5]. The knowledge descriptions carry discoverable affordances and interpretable signifiers, which can then be resolved by agents [3, 35]. This approach might demand extended web standards that make the web a suitable environment for software agents. It might also require introducing architectural constraints that web platforms must adhere to in developing and changing their platforms, such as providing a knowledge description where bots can read descriptions of their affordances.

The second category of approaches focuses on using various behavioral cloning and reinforcement learning techniques [22]. One example is by Shi et al. [47], where they introduce a simulation and live training environment to enable bots to complete web interaction activities utilizing keyboard and mouse actions. Recent efforts extend these approaches by leveraging large language models (LLMs) for web page understanding and autonomous web navigation [21, 25]. The results from both techniques and similar approaches reveal the size of the gap between human users and bots [47, 21]. Both approaches still need to solve the problem of variability and generalizability of perception and action. Though approaches that leverage the hypermedia knowledge of platforms with affordance and signifier descriptions could serve as placeholders, real bots with generalizable capabilities would need more autonomous models yet.

Besides this, some design assumptions consider the environment and the bot as one. As a result, they may attempt to design agents as an integrated part of the platforms or try to "botify" and "cognify" or orient web services as agents. However, alternatively, the whole notion of a *user-like* bot inherently considers the bot to have an autonomous presence separate from the web platforms it accesses. Fig. 1 illustrates the basic perspective in a vertically separate design and development of the bot and the web platforms it operates in. This strict separation enables both the environment and the bot to evolve independently.
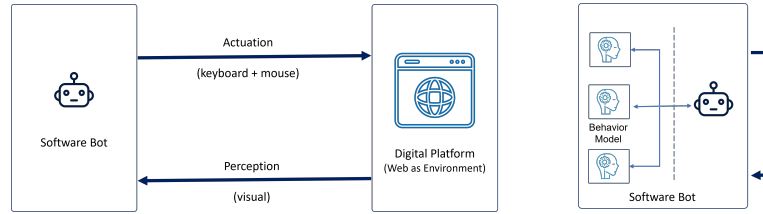
Fig. 1. A decoupled bot-environment viewpoint.

Fig. 2. Separate bot and behavior models.

3   The Behavior Problem

Most user activities on digital platforms are complex behaviors resulting from human users' underlining intentions, goals, and belief systems. Although a bot operating in digital spaces need not fully emulate humans to achieve generalizable behavior, it is essential to consider the intricacies and sophistication of human users' behavior on the web as an example [44]. To that end, engineering bots with similar behavior models might take account of existing approaches to measuring generalizable user behavior while not having to replicate human cognition as such [51].
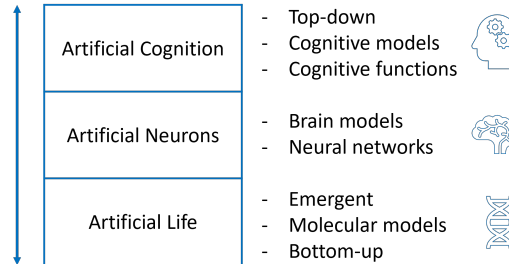
Fig. 3. The abstraction ladder in modeling machine intelligence.

Current models for engineering intelligent behavior come from three prospective categories of approaches. Each approach takes natural or human intelligence as its inspiration and models it at different levels of abstraction. The three methods differ mainly in how they try to understand intelligence and where they start the abstraction for modeling intelligence. Fig. 3 illustrates this ladder of abstraction in modeling machine intelligence. The abstractions start either at artificial cognition, artificial neurons, or artificial life or consciousness [18, 49, 49]. These relate to models and techniques for each approach to enacting intelligent behavior based, respectively, on high-level cognitive functions, artificial neural networks (ANNs), or more physical and bottom-up approaches starting

at molecular or atomic levels.

Artificial Cognition: in cognitive modeling, efforts to model cognition are inspired by the brain's high-level cognitive functions, such as memory. Abstraction is at the topmost level compared to the other two. Most assumptions come from studies and understanding of the cognitive sciences. Cognitive models use diverse techniques, such as production rules, dynamical systems, and quantum models, to model particular cognitive capabilities [46]. Although cognitive models use methods from other approaches, such as ANNs, they do not necessarily adhere to underlying mechanisms in the brain [53, 18]. Promising experimental research examples that heavily rely on artificial cognitive models, aka cognitive architectures, are works such as the OpenCog (Hyperon) and the iCub project [18, 53].

Artificial Neurons: artificial neurons as brain models aim to understand, model, and simulate underlying computational mechanisms and functions based on assumptions and studies from the neurosciences [11]. Discoveries from neuroscience are utilized to drive brain-based computational principles. Sometimes, these approaches are referred to as *Brain-derived AI* or *NeuroAI* models [55, 32, 39]. Due to the attention given to the underline principles of computation in the brain, they strictly differ from the brain-inspired cognitive models. Practices are mainly advancements in artificial neural networks, such as deep learning. Large-scale brain simulation research and new hardware in neuromorphic computing, such as *SpiNNaker* and *Loihi*, also contribute to research efforts in this area. Computational capabilities in neuromorphic computing enable particular types of neural networks closer to brain computational principles, such as *Spiking neural networks* [37]. Brain-derived AI approaches, with neurorobotics, aim to achieve embodiment using fully developed morphologies, either physical or virtual models. The Neurorobotics Platform (NRP) is an example of such efforts to develop and simulate embodied systems. The NRP is a neurorobotics simulation environment to connect simulated brains to simulated bodies of robots [31, 30].

Artificial Life (aLife) Attempts to model consciousness, particularly in the case of artificial life, start with a bottom-up approach at a physical or molecular level [49]. Most synthesizing efforts to models of intelligence in artificial life are simulations with digital avatars.

However, in the context of bots on web platforms, employing similar integrated behavior models to those mentioned above is still a challenge. Thus, in addition to the proposed separation of the bot and environment, decoupling a bot's basic skeleton and behavior models is architecturally significant. Fig. 2 illustrates the separate structure of a bot and its behavior models. The vertical separation allows behavior models and bot skeletons to change independently, maintaining the possibility of dynamic coupling.

4   Bot Communication Challenges

In Multi-Agent Systems (MAS), agent-to-agent communication heavily relies on agent communication languages (ACLs) such as FIPA-ACL, standardized

by the Foundation for Intelligent Physical Agents(FIPA) consortium [48, 57, 26, 23]. However, in mixed reality environments, where bots and humans share and collaborate in digital space and beliefs, communication cannot rely only on ACLs and APIs [24].

To that end, a cognitive bot with artificial general intelligence (AGI) must possess communications capabilities to address humans and software agents with diverse communication skills. Communication capabilities should include diverse possibilities like email, dialogue systems, voice, blogging, and micro-blogging.

Large language models (LLMs) have recently shown significant progress in natural language processing and visual perception that could be utilized for bot and human communication [25, 21].

5   Integration and Cooperation Challenges

Researchers assert that the grand challenge in AGI remains in integrating different intelligence components to enable the emergence of advanced generalizable behavior or even collective intelligence [38, 18, 50, 13]. The intelligence solutions to integrate include learning, memory, perception, actuation, and other cognitive capabilities [16]. Theories and assumptions developed by proponents include approaches based on cognitive synergy, the free energy principle, and integrated information theory [17, 13, 50].

In practice, however, integration and cooperation of software agents are implemented mainly by utilizing methods such as ontologies, APIs, message routing, communication protocols, and middleware like the blackboard pattern [8, 57, 1].

From a software engineering perspective, basic architectural requirements for the context of bots operating on digital platforms are possibilities for the evolvability of bots into collective understanding with shared beliefs, stigmergy, or sharing common behavior models to learn and evolve. Other minor concerns are the hosting and distribution of bots and their behavior models, which could be on a cloud or individual nodes.

Fig. 4 shows a simple diagram representing the integrated parts, i.e., bots, shared behavior models, and the environment. B represents the possible number of bots. BM represents the shared behavior models. E represents the web environment and its variability. The lines represent communication channels. H denotes the human users that participate and share the digital space.

6   Bot Ethics and Trust

Concerns and challenges in AGI are diverse. They touch on various aspects of society and politics. One such case is the impact, real-world implications, of such user-like bots on privacy, security, ethics, and trust with humans [20, 4, 6, 7].

User-like bots, emulating human users' perceptual and interaction techniques, can easily pass bot detection tests and risk exploitation for malicious use cases
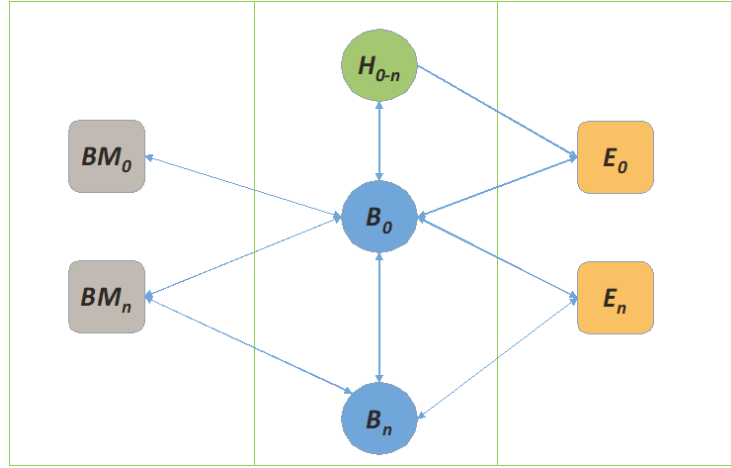
Fig. 4. Representation of integrated parts, i.e., bots, shared behavior models, and the environment.

to deceive and attack web platforms. They could also extend their perceptual capabilities beyond the web with connected devices, such as microphones and cameras, affecting the personal privacy of others. Possible threats include spamming, cyberattacks to overwhelm platforms, and even unfair use of web platforms, for surveillance or illicit financial gains. In WoT context, for instance, bots could affect the real world, such as smart factories and automated services, compromising physical devices and processes with significant security implications [29].

Hypothetically, intelligent social bots could share their beliefs on social platforms as any human user or beyond with superb reasoning and argumentation skills. These cases could negatively impact society by exposing people and software agents to unexpected, misaligned norms and moral beliefs. Furthermore, deploying advanced cognitive bots as digital workforces may result in unforeseen negative economic consequences. Short-term issues could include unemployment, while long-term concerns may involve ethical dilemmas surrounding bot ownership rights, bot farming, or 'enslavement' [20]. Accordingly, these ethical concerns may affect the legality of cognitive bot development, potentially impeding their engineering and deployment. Alternatively, this could introduce new legal aspects regarding regulation, standards, and ethics for integrating and governing bots within emerging socio-technical ecosystems [29].

Despite these concerns, software agents' current and promising applications can positively impact numerous aspects of society. Cognitive automation, for example, is driving increased demand for cognitive agents in Industry 4.0, digital twins, and other digital environments [9, 54, 28]. Early implementations, like Wikipedia bots, already play a significant role in fact-checking and other knowledge-based tasks. On platforms like GitHub, bots assist and automate

development tasks [56, 27]. Future cognitive bots could also benefit society in advancing scientific research, like medical research, by participating in knowledge processing and providing valuable new scientific insights, which significantly outweigh their potential risks.

Today, digital platforms handle simple crawling and API-based bots with crawling policies and controlled exposure of APIs. However, advanced user-like bots like the ones envisioned in this report will require more complex mechanisms to govern and control behavior and belief-sharing [20, 29]. One approach towards this, ethics and trust by design, emphasizes providing protocols and policies to be followed by developers and engineering organizations to incorporate trust models and ethical frameworks at design and architectural stages [20]. Another approach proposes norms and user policies with penalties for agents to acknowledge, understand, and adhere to, similar to what human users would do on digital platforms [29]. In return, norm and value-aware bots could establish participation, trust, and compliance while facing the consequences of noncompliance. They may also contribute to the collective value and norm revisions or creations becoming part of possible viable socio-technical ecosystems [29, 43].

However, ensuring safety and trust in such ecosystems will require diverse approaches. In addition to the machine-readable norms and policies targeting cognitive agents, tackling ethical and trust issues with transparent and explainable design and engineering processes is also essential. As a result, research on developing advanced cognitive bots must also address critical challenges in engineering trustworthy, secure, and verifiable AGI bots within the design and architecture [7].

7   Conclusion

The study presented architectural research challenges toward designing and developing a new line of user-like cognitive bots operating autonomously on digital platforms. Key challenges, such as the transduction problem, are discussed in the context of digital web platforms' access, user-like visual interaction, and autonomous navigation. Bot-environment separation for autonomy and bot skeleton and behavior model separation for better evolvability are also discussed as architectural recommendations.

Furthermore, challenges in enacting generalizable behavior, communication, and cooperation are also presented. Finally, challenges on cognitive bots' ethics, trust, implications, and future impacts are also discussed as part of architectural concerns.

As an outlook, a good starting point for future work would be to conceptualize a detailed implementation architecture and construct a software bot by utilizing existing cognitive models. These systems can demonstrate the concept and allow further detailed analysis through empirical data and benchmarks.

References

1. Boissier, O., Ciortea, A., Harth, A., Ricci, A.: Autonomous agents on the web. In: DS 21072. p. 100p (2021)
2. Brooks, R.A.: Intelligence without representation. AI 47(1-3), 139–159 (1991)
3. Charpenay, V., Amundsen, M., Mayer, S., Padget, J., Schraudner, D., Vachtsevanou, D.: 6.2 affordances and signifiers. Autonomous Agents on the Web p. 67
4. Christian, B.: The Alignment Problem: Machine Learning and Human Values. WW Norton & Company (2020)
5. Ciortea, A., Mayer, S., Boissier, O., Gandon, F.: Exploiting interaction affordances: on engineering autonomous systems for the web of things (2019)
6. Coeckelbergh, M.: AI Ethics. Mit Press (2020)
7. Dengel, A., Etzioni, O., DeCario, N., Hoos, H., Li, F.F., Tsujii, J., Traverso, P.: Next big challenges in core ai technology. Reflections on AI for Humanity pp. 90–115 (2021)
8. Dorri, A., Kanhere, S.S., Jurdak, R.: Multi-agent systems: A survey. IEEE Access 6, 28573–28593 (2018)
9. Engel, C., Ebel, P., Leimeister, J.M.: Cognitive automation. Electronic Markets 32(1), 339–350 (2022)
10. Etzioni, O.: Intelligence without robots: A reply to brooks. AI Magazine 14(4), 7–7 (1993)
11. Fan, X., Markram, H.: A brief history of simulation neuroscience. Frontiers in neuroinformatics 13, 32 (2019)
12. Fountain, A.M., Hall, W., Heath, I., Davis, H.C.: Microcosm: An open model for hypermedia with dynamic linking. In: ECHT. vol. 90, pp. 298–311 (1990)
13. Friston, K.: The free-energy principle: a unified brain theory? Nature reviews neuroscience 11(2), 127–138 (2010)
14. Gallwitz, F., Kreil, M.: The rise and fall of'social bot'research. SSRN (2021)
15. Gibson, J.J.: The theory of affordances. Hilldale, USA 1(2), 67–82 (1977)
16. Goertzel, B.: Artificial general intelligence: concept, state of the art, and future prospects. AGI 5(1), 1 (2014)
17. Goertzel, B.: Toward a formal model of cognitive synergy. arXiv:1703.04361 (2017)
18. Goertzel, B., Pennachin, C., Geisweiller, N.: Engineering general intelligence, part 1. Atlantis Thinking Machines 5 (2014)
19. Goertzel, B., Pennachin, C., Geisweiller, N.: Engineering General Intelligence, Part 2, vol. 6. Springer (2014)
20. Goertzel, B., Pennachin, C., Geisweiller, N., Goertzel, B., Pennachin, C., Geisweiller, N.: The engineering and development of ethics. Engineering General Intelligence, Part 1 pp. 245–287 (2014)
21. Gur, I., Nachum, O., Miao, Y., Safdari, M., Huang, A., Chowdhery, A., Narang, S., Fiedel, N., Faust, A.: Understanding html with large language models. arXiv:2210.03945 (2022)
22. Gur, I., Rueckert, U., Faust, A., Hakkani-Tur, D.: Learning to navigate the web. arXiv:1812.09195 (2018)
23. Hillmann, P., Uhlig, T., Rodosek, G.D., Rose, O.: A novel multi-agent system for complex scheduling problems. In: WSC'14. pp. 231–241. IEEE (2014)
24. Holz, T., Campbell, A.G., O'Hare, G.M., Stafford, J.W., Martin, A., Dragone, M.: Mira—mixed reality agents. Human-computer studies 69(4), 251–268 (2011)
25. Huang, S., Dong, L., Wang, W., Hao, Y., Singhal, S., Ma, S., Lv, T., Cui, L., Mohammed, O.K., Liu, Q., et al.: Language is not all you need: Aligning perception with language models. arXiv:2302.14045 (2023)

26. Hübner, J.F.: 4.5 about the place of agents in the web. Autonomous Agents on the Web p. 44
27. Hukal, P., Berente, N., Germonprez, M., Schecter, A.: Bots coordinating work in open source software projects. Computer 52(9), 52–60 (2019)
28. Ivančić, L., Suša Vugec, D., Bosilj Vukšić, V.: Robotic process automation: systematic literature review. In: BPM. pp. 280–295. Springer (2019)
29. Kampik, T., Boissier, O., Kirrane, S., Mansour, A., Padget, J., Payne, T., Singh, M., Tamma, V., Zimmermann, A.: Norms and Policies for Agents on the Web (2021)
30. Knoll, A., Gewaltig, M.O., Sanders, J., Oberst, J.: Neurorobotics: A strategic pillar of the human brain project. Science Robotics pp. 2–3 (2016)
31. Knoll, A., Röhrbein, F., Kuhn, A., Akl, M., Sharma, K.: Neurorobotics. Informatik-Spektrum 40(2), 161–164 (2017)
32. Knoll, A., Walter, F.: Neurorobotics–a unique opportunity for ground breaking research. Tech. rep., Chair of Robotics, AI and Real-Time Systems (2019)
33. Lebeuf, C., Storey, M.A., Zagalsky, A.: Software bots. IEEE Software 35(1), 18–23 (2017)
34. Lebeuf, C.R.: A Taxonomy of Software Bots: Towards a deeper understanding of Software Bot Characteristics. Ph.D. thesis (2018)
35. Lemee, J., Vachtsevanou, D., Mayer, S., Ciortea, A.: Signifiers for affordance-driven multi-agent systems. In: International Workshop on Engineering Multi-Agent Systems (2022)
36. Lieto, A., Bhatt, M., Oltramari, A., Vernon, D.: The role of cognitive architectures in general artificial intelligence (2018)
37. Maass, W.: Networks of spiking neurons: the third generation of neural network models. Neural networks 10(9), 1659–1671 (1997)
38. Minsky, M.: Society of mind. Simon and Schuster (1988)
39. Momennejad, I.: A rubric for human-like agents and neuroai. Philosophical Transactions of the Royal Society B 378(1869), 20210446 (2023)
40. Nelson, T.H.: Complex information processing: a file structure for the complex, the changing and the indeterminate. In: Proceedings of the 1965 20th national conference. pp. 84–100 (1965)
41. Norvig, P.R., Intelligence, S.A.: A modern approach. Prentice Hall Upper Saddle River, NJ, USA: Rani, M., Nayak, R., & Vyas, OP (2015). An ontology-based adaptive personalized e-learning system, assisted by software agents on cloud storage. Knowledge-Based Systems 90, 33–48 (2002)
42. Nye, B.D., Silverman, B.G.: Affordances in ai (2012)
43. Patrick, A.S.: Building trustworthy software agents. IEEE Internet Computing 6(6), 46–53 (2002)
44. Pennachin, C., Goertzel, B.: Contemporary approaches to artificial general intelligence. AGI pp. 1–30 (2007)
45. Rosenbloom, P.S.: Thoughts on architecture. In: AGI. pp. 364–373. Springer (2023)
46. Schöner, G.: Dynamical systems approaches to cognition. Cambridge handbook of computational cognitive modeling pp. 101–126 (2008)
47. Shi, T., Karpathy, A., Fan, L., Hernandez, J., Liang, P.: World of bits: An open-domain platform for web-based agents. In: ICML. pp. 3135–3144. PMLR (2017)
48. Soon, G.K., On, C.K., Anthony, P., Hamdan, A.R.: A review on agent communication language. ICCST pp. 481–491 (2019)
49. Taylor, T., Auerbach, J.E., Bongard, J., Clune, J., Hickinbotham, S., Ofria, C., Oka, M., Risi, S., Stanley, K.O., Yosinski, J.: WebAL comes of age: A review of the first 21 years of artificial life on the web. Artificial life 22(3), 364–407 (2016)

50. Tononi, G.: Integrated information theory. Scholarpedia 10(1), 4164 (2015)
51. Turing, A.: Computing machinery and intelligence (1950)
52. Vachtsevanou, D., Ciortea, A., Mayer, S., Lemée, J.: Signifiers as a first-class abstraction in hypermedia multi-agent systems. arXiv:2302.06970 (2023)
53. Vernon, D.: Cognitive Architectures. In: Cognitive Robotics. The MIT Press (2022)
54. Vogel-Heuser, B., Seitz, M., Salazar, L.A.C., Gehlhoff, F., Dogan, A., Fay, A.: Multi-agent systems to enable industry 4.0. at-Automatisierungstechnik 68(6), 445–458 (2020)
55. Walter, F.: Advanced Embodied Learning. Ph.D. thesis, Technische Universität München (2021)
56. Wessel, M., De Souza, B.M., Steinmacher, I., Wiese, I.S., Polato, I., Chaves, A.P., Gerosa, M.A.: The power of bots: Characterizing and understanding bots in oss projects. PACM:HCI 2(CSCW), 1–19 (2018)
57. Wooldridge, M.: An introduction to multiagent systems. John wiley & sons (2009)
58. Yampolskiy, R.V.: AI-complete, aI-hard, or aI-easy–classification of problems in AI. In: MAICS (2012)