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

Abstract. Software bots operating in multiple virtual digital platforms must understand the platforms' affordances and behave like human users. Platform affordances or features differ from one application platform to another or through a life cycle, requiring such bots to be adaptable. Moreover, bots in such platforms could cooperate with humans or other software agents for work or to learn specific behavior patterns. 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 by directing attention to significant architectural research challenges in developing cognitive bots endowed with complex behavior for operation 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 [1,2]. An example scenario would be a "*user-like*" bot that could access web platforms 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 [3,4]. The bot would recognize and understand the differences and variability between different environments' affordances. If a platform or service has extensions to physical bodies or devices, as in the Web of Things (WoT), it would also have control of or possibilities to interact with an outer web or service application world.

Ideally, a bot could also be independent of a specific platform. A user-like social bot, for instance, would be able to recognize and understand

P. Hammer et al. (Eds.): AGI 2023, LNAI 13921, pp. 105–114, 2023.
https://doi.org/10.1007/978-3-031-33469-6_11

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 social media environment, such as when hypermedia interactivity features and functionalities change. Such a bot could also learn and develop to derive its goals and intentions from these digital microenvironments and take goal-directed targeted action to achieve them [5]. 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 desiderata 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. Instead, most bots are either conversational interfaces or question-and-answer knowledge agents [1]. Others only perform automated repetitive tasks based on pre-given rules, lacking autonomy and other advanced cognitive skills [6,7]. The problems of realizing these desiderata are, therefore, complex and challenging [8,9]. Solutions must address different areas, such as transduction and autonomous action, to achieve advanced generalizable intelligent behavior [10,11].

Problems spanning diverse domains require architectural solutions. Accordingly, these challenges also necessitate that researchers address the structural and dynamic elements of such systems from an architectural perspective. [12–14]. For this reason, this paper outlines the architectural research agendas to address the challenges in 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 developing advanced cognitive models in software bots. Section 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 distinct microenvironments within digital microcosms [15]. They offer a microhabitat for their users' diverse digital experiences. These experiences mainly transpire from the elements of interaction and action, or the hypermedia, within web environments [15,16]. Hypermedia connects and extends the user experience, linking to further dimensions of the web-worlds, which means more pages and interactive elements from the user's perspective. The interaction elements are considered affordances in the digital space [3,4], analogous to the biological concept of affordances from environmental psychology [17]. Affordances can also be accompanied by signifiers. Signifiers reveal or indicate possibilities for actions associated with affordances [4,18]. 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 this web environment, its content, and its 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 [19,20].

Today, there are different approaches to this problem. The first category of approaches provides knowledge about the environment for different levels of observability using APIs or knowledge descriptions. With API-based approaches, developers create bots 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 WoT, attempt to address this challenge by using knowledge models and standards that could enable agents to perceive the web by exposing hypermedia affordances and signifiers [3,21]. The knowledge descriptions carry discoverable affordances and interpretable signifiers, which can then be resolved by agents [3,4]. 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 uses various behavioral cloning and reinforcement learning techniques [22]. One example is by Shi et al. [23], 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 [24,25]. The results from both techniques and similar approaches reveal the size of the gap between human users and bots [23,24].

Both approach categories still need to solve the problem of variability and generalizability of perception and action. Approaches that leverage the hypermedia knowledge of platforms with affordance and signifier descriptions could serve as placeholders, but 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. Alternatively, the whole notion of a *user-like* bot inherently assumes the bot to have an autonomous presence separate from the web platforms it accesses. Figure 1 illustrates the basic perspective in a vertically separate design, the bot, and the web platforms it operates in. This strict separation enables both the environment and the bot to evolve independently.

3 The Behavior Problem

Most user activities on digital platforms are complex behaviors resulting from human users’ underlying intentions, goals, and belief systems. Although a bot

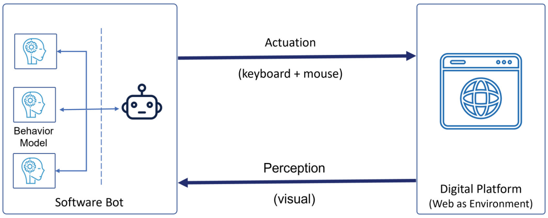

Fig. 1. A decoupled bot-environment and bot-behavior (*left*) viewpoint.

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 during bot design [26]. To that end, engineering bots with behavior models similar to human users might take into account existing approaches of measuring generalizable user behavior while not having to replicate human cognition as such [27].

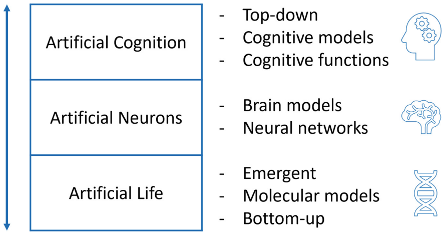

Fig. 2. The abstraction ladder in modeling machine intelligence.

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. Figure 2 illustrates this ladder of abstraction in modeling machine intelligence. The abstractions start either at artificial cognition, artificial neurons, or artificial life or consciousness [10,28]. These abstractions aim to enact 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. Most assumptions are based on studies and understandings in the cognitive sciences. Cognitive models use diverse techniques such as production rules, dynamical systems, and quantum models to model particular cognitive capabilities [29]. Although cognitive models use methods from other approaches, such as ANNs, they do not necessarily adhere to underlying mechanisms in the brain [10,30]. Works such as the OpenCog (Hyperon) and the iCub project are promising experimental research examples that heavily rely on artificial cognitive models, i.e., cognitive architectures [10,30].

Artificial Neurons: brain models which use artificial neurons aim to understand, model, and simulate underlying computational mechanisms and functions based on assumptions and studies in neuroscience [31]. Discoveries from neuroscience are utilized to derive brain-based computational principles. Sometimes, these approaches are referred to as *Brain-derived AI* or *NeuroAI* models [32–34]. Due to the attention given to the underlying principles of computation in the brain, they strictly differ from the brain-inspired cognitive models. Applications of these models are mainly advancements in artificial neural networks, such as deep learning. Large-scale brain simulation research and new hardware development in neuromorphic computing, such as *SpiNNaker* and *Loihi*, also contribute to research efforts in this area. Some neuromorphic hardware enables close adherence to brain computational principles in particular types of neural networks, such as *Spiking neural networks* [32,35]. Brain-derived AI approaches with neurorobotics aim to achieve embodiment using fully developed morphologies, which are either physical or virtual. The Neurorobotics Platform (NRP) is an example of such efforts to develop and simulate embodied systems. The NRP is a neurorobotics simulation environment which connects simulated brains to simulated bodies of robots [36].

Artificial Life (aLife): aLife attempts to model consciousness. To do this, researchers and developers start with a bottom-up approach at a physical or molecular level [28]. Most synthesizing efforts to model intelligence in artificial life are simulations with digital avatars.

In the context of bots on web platforms, employing integrated behavior models, such as the NRP and OpenCog 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 important. Figure 1, *left*, illustrates the separate structure of a bot and its behavior models. The bot's core skeleton, for example, might have sensory and interaction elements as virtual actuators that enable its operation using the keyboard and mouse actions. 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 [19,37–39]. However, in mixed reality environments, where bots and humans share and collaborate in digital spaces, communication cannot rely only on ACLs and APIs [40].

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 [24,25].

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 [10,41–43]. The intelligence solutions to integrate include learning, memory, perception, actuation, and other cognitive capabilities [44]. Theories and assumptions developed by proponents include approaches based on cognitive synergy, the free energy principle, and integrated information theory [5,42,43].

In practice, however, integration and cooperation of bots are implemented mainly by utilizing methods such as ontologies, APIs, message routing, communication protocols, and middleware like the blackboard pattern [19,45,46].

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, transfer learned experience, and evolve. Other concerns are the hosting, which could be on a cloud or individual nodes, scaling, and distribution of bots and their behavior models.

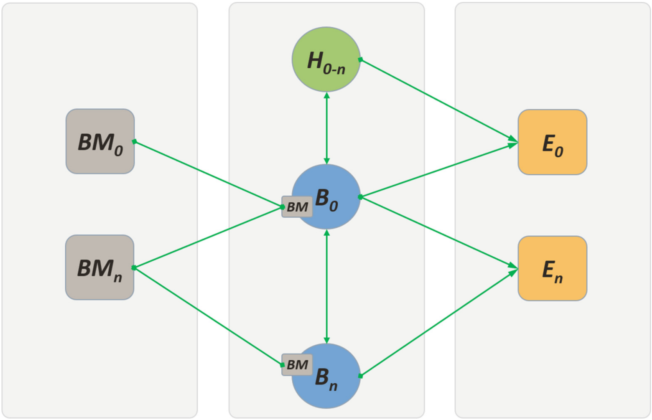

Fig. 3. Representation of integrated parts, i.e., bots, shared behavior models, and the web environments.

Figure 3 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 and individual 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 and have real-world implications, such as the impact of user-like bots on privacy, security, ethics, and trust with humans [47–49]. User-like bots, emulating human users' perceptual and interaction techniques, can easily pass bot detection tests and risk exploitation for malicious use cases 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 smart factories and automated services in the real world, compromising physical devices and processes with significant security implications [50].

Hypothetically, intelligent social bots could share their beliefs on social platforms similar to or better than any human user, 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' [47]. 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 [50].

Despite these concerns, bots' current and potential 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 [6,7,51]. 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 [52]. Future cognitive bots could also benefit society by participating in knowledge processing and providing valuable new scientific insights, such as medical advancements, 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 their behavior and belief-sharing [47,50]. One approach towards this is ethics and trust by design, which recommends protocols and policies for developers and engineering organizations to incorporate trust models and ethical frameworks at the design and architectural stages [47]. 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 [50,53]. In return, norm and value-aware bots could establish participation, trust, and compliance while facing the consequences of noncompliance. They may also contribute to revising and creating collective values and norms, possibly becoming part of viable socio-technical ecosystems [50,54].

However, ensuring safety and trust in such ecosystems will require diverse approaches. In addition to providing machine-readable norms and policies targeting cognitive agents, it is essential to tackle ethical and trust issues with transparent and explainable design and engineering processes at each stage. For instance, the European Union (EU) recommends a three-phase human intervention approach at the design phase, at the development and training phase, and at runtime with oversight and override possibilities [55]. As a result, research on developing advanced cognitive bots must also address critical challenges in engineering trustworthy, secure, and verifiable AGI bots employing hybrid approaches.

7 Conclusion

The study presented architectural research challenges in 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. In the architecture, we recommend bot-environment separation to realize bot autonomy and bot skeleton and behavior model separation for better evolvability. Also, bot communication capabilities should include diverse possibilities like email, dialogue systems, and blogging. We recommend utilizing shared behavior models for transfer learning or collective intelligence to enact generalizable behavior. Finally, we discussed cognitive bots' ethical implications and potential long-term effects, proposing to adopt hybrid approaches that incorporate these aspects into the architecture and the life cycle of bots.

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

References

1. Lebeuf, C.R.: A taxonomy of software bots: towards a deeper understanding of software bot characteristics. Ph.D. thesis, UVic (2018)
2. Etzioni, O.: Intelligence without robots: a reply to brooks. AI Mag. **14**(4), 7 (1993)

3. Charpenay, V., Amundsen, M., Mayer, S., Padget, J., Schraudner, D., Vachtsevanou, D.: 6.2 affordances and signifiers. In: Autonomous Agents on the Web, p. 67 (2021)
4. Lemee, J., Vachtsevanou, D., Mayer, S., Ciortea, A.: Signifiers for affordance-driven multi-agent systems. In: EMAS (2022)
5. Goertzel, B.: Toward a formal model of cognitive synergy. In: AGI (2017)
6. Ivančić, L., Suša Vugec, D., Bosilj Vukšić, V.: Robotic process automation: systematic literature review. In: Di Ciccio, C., et al. (eds.) BPM 2019. LNBIP, vol. 361, pp. 280–295. Springer, Cham (2019). https://doi.org/10.1007/978-3-030-30429-4_19
7. Engel, C., Ebel, P., Leimeister, J.M.: Cognitive automation. Electron. Markets **32**(1), 339–350 (2022). https://doi.org/10.1007/s12525-021-00519-7
8. Russell, S.J.: Artificial Intelligence a Modern Approach. Pearson Education (2010)
9. Yampolskiy, R.V.: AI-complete, AI-hard, or AI-easy-classification of problems in AI. In: MAICS (2012)
10. Goertzel, B., Pennachin, C., Geisweiller, N.: Engineering General Intelligence, Part 1. Atlantis Thinking Machines, vol. 5. Atlantis Press, Paris (2014)
11. Vernon, D., von Hofsten, C., Fadiga, L.: Desiderata for developmental cognitive architectures. BICA **18**, 116–127 (2016)
12. Goertzel, B., Pennachin, C., Geisweiller, N.: Engineering General Intelligence, Part 2, vol. 6. Springer, Heidelberg (2014). https://doi.org/10.2991/978-94-6239-030-0
13. Rosenbloom, P.S.: Thoughts on architecture. In: Goertzel, B., Iklé, M., Potapov, A., Ponomaryov, D. (eds.) AGI. MNCS, vol. 13539, pp. 364–373. Springer, Cham (2023). https://doi.org/10.1007/978-3-031-19907-3_35
14. Lieto, A., Bhatt, M., Oltramari, A., Vernon, D.: The role of cognitive architectures in general artificial intelligence. Cogn. Syst. Res. **48**, 1–3 (2018)
15. Fountain, A.M., Hall, W., Heath, I., Davis, H.C.: MICROCOSM: an open model for hypermedia with dynamic linking. In: ECHT (1990)
16. Nelson, T.H.: Complex information processing: a file structure for the complex, the changing and the indeterminate. In: ACM National Conference (1965)
17. Gibson, J.J.: The theory of affordances, Hilldale, USA, vol. 1, no. 2, pp. 67–82 (1977)
18. Vachtsevanou, D., Ciortea, A., Mayer, S., Lemée, J.: Signifiers as a first-class abstraction in hypermedia multi-agent systems. In: EKAW (2023)
19. Wooldridge, M.: An Introduction to Multiagent Systems. Wiley, Hoboken (2009)
20. Brooks, R.A.: Intelligence without representation. AI **47**(1–3), 139–159 (1991)
21. Ciortea, A., Mayer, S., Boissier, O., Gandon, F.: Exploiting interaction affordances: on engineering autonomous systems for the web of things (2019)
22. Gur, I., Rueckert, U., Faust, A., Hakkani-Tur, D.: Learning to navigate the web. arXiv (2018)
23. 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)
24. Gur, I., et al.: Understanding HTML with large language models. arXiv (2022)
25. Huang, S., et al.: Language is not all you need: aligning perception with language models. arXiv (2023)
26. Pennachin, C., Goertzel, B.: Contemporary approaches to artificial general intelligence. In: Goertzel, B., Pennachin, C. (eds.) AGI. COGTECH, pp. 1–30. Springer, Heidelberg (2007). https://doi.org/10.1007/978-3-540-68677-4_1
27. Computing machinery and intelligence-AM turing (1950)
28. Taylor, T., et al.: WebAL comes of age: a review of the first 21 years of artificial life on the web. Artif. Life **22**(3), 364–407 (2016)

29. Schöner, G.: Dynamical systems approaches to cognition. In: Cambridge Handbook of Computational Cognitive Modeling, pp. 101–126 (2008)
30. Vernon, D.: Cognitive architectures. In: Cognitive Robotics. MIT Press (2022)
31. Fan, X., Markram, H.: A brief history of simulation neuroscience. Front. Neuroinform. **13**, 32 (2019)
32. Walter, F.: Advanced embodied learning. Ph.D. thesis, TUM (2021)
33. Knoll, A., Walter, F.: Neurorobotics-a unique opportunity for ground breaking research (2019)
34. Momennejad, I.: A rubric for human-like agents and NeuroAI. Philos. Trans. **378**(1869), 20210446 (2023)
35. Maass, W.: Networks of spiking neurons: the third generation of neural network models. Neural Netw. **10**(9), 1659–1671 (1997)
36. Knoll, A., Gewaltig, M.-O., Sanders, J., Oberst, J.: Neurorobotics: a strategic pillar of the human brain project. Sci. Robot. 2–3 (2016)
37. Soon, G.K., On, C.K., Anthony, P., Hamdan, A.R.: A review on agent communication language. In: ICCST, pp. 481–491 (2019)
38. Hübner, J.F.: 4.5 about the place of agents in the web. In: Autonomous Agents on the Web, p. 44 (2021)
39. Hillmann, P., Uhlig, T., Rodosek, G.D., Rose, O.: A novel multi-agent system for complex scheduling problems. In: WSC 2014. IEEE (2014)
40. Holz, T., Campbell, A.G., O'Hare, G.M.P., Stafford, J.W., Martin, A., Dragone, M.: Mira-mixed reality agents. IJHC **69**(4), 251–268 (2011)
41. Minsky, M.: Society of Mind. Simon and Schuster (1988)
42. Tononi, G.: Integrated information theory. Scholarpedia **10**(1), 4164 (2015)
43. Friston, K.: The free-energy principle: a unified brain theory? Nat. Rev. Neurosci. **11**(2), 127–138 (2010)
44. Goertzel, B.: Artificial general intelligence: concept, state of the art, and future prospects. AGI **5**(1), 1 (2014)
45. Dorri, A., Kanhere, S.S., Jurdak, R.: Multi-agent systems: a survey. IEEE Access **6**, 28573–28593 (2018)
46. Boissier, O., Ciortea, A., Harth, A., Ricci, A.: Autonomous agents on the web. In: DS 21072, p. 100p (2021)
47. Goertzel, B., Pennachin, C., Geisweiller, N., Goertzel, B., Pennachin, C., Geisweiller, N.: The engineering and development of ethics. In: Engineering General Intelligence, Part 1, pp. 245–287 (2014)
48. Christian, B.: The Alignment Problem: Machine Learning and Human Values. WW Norton & Company (2020)
49. Coeckelbergh, M.: AI Ethics. MIT Press, Cambridge (2020)
50. Kampik, T., et al.: Norms and policies for agents on the web. In: Autonomous Agents on the Web (2021)
51. Vogel-Heuser, B., Seitz, M., Cruz Salazar, L.A., Gehlhoff, F., Dogan, A., Fay, A.: Multi-agent systems to enable industry 4.0. at-Automatisierungstechnik **68**(6), 445–458 (2020)
52. Hukal, P., Berente, N., Germonprez, M., Schecter, A.: Bots coordinating work in open source software projects. Computer **52**(9), 52–60 (2019)
53. Kampik, T., et al.: Governance of autonomous agents on the web: challenges and opportunities. ACM TOIT **22**(4), 1–31 (2022)
54. Patrick, A.S.: Building trustworthy software agents. IEEE IC **6**(6), 46–53 (2002)
55. Kaur, D., Uslu, S., Rittichier, K.J., Durresi, A.: Trustworthy artificial intelligence: a review. ACM CSUR **55**(2), 1–38 (2022)