Cognitive Planning for Persuasive Multimodal Interaction

Emiliano Lorini[1], Magalie Ochs[2], and Nicolas Sabouret[3]

[1] IRIT, CNRS, Toulouse University, France
[2] LIS, Aix-Marseille University, France
[3] LISN, CNRS, Université Paris-Saclay, France

Abstract. This paper proposes guidelines for the design of embodied social agents endowed with cognitive planning capabilities to be used in the context of persuasive multimodal interaction. Cognitive planning is the kind of planning aimed at finding and then executing an informative plan with the goal of influencing the cognitive state or the behavior of the interlocutor.

1 Introduction

An agent's cognitive state encompasses its epistemic attitudes (e.g., beliefs and opinions), motivational attitudes (e.g., desires, moral values, preferences and intentions) and emotions. Their causal relationships as well as their causal influence on the agent's behavior are objects of study in cognitive psychology [1] and philosophy of mind [32]. Figure 1 provides a schematic representation of these causal relationships. For instance,

- the agent's preferences causally depend on both its endogenous intrinsic motivations (*alias* desires) and ethical values;
- the agent's intentions are the output of its decision-making process and causally depend on its beliefs and preferences;
- the agent's emotions are triggered by its beliefs, desires and moral concerns;
- the agent's emotions may affect its behavior by bypassing its deliberation process (e.g., fear triggered by the perception of a danger can cause an automatic response of escape).

Cognitive planning consists in an agent (the influencer) trying to find and then execute a plan aimed at changing the cognitive state or the behavior of another agent (the influenced or target agent). The concept of cognitive planning lies at the intersection between AI, cognitive sciences, social psychology and ethics. Thus, to be properly understood, it has to be investigated from an interdisciplinary perspective. From a human science point of view, cognitive planning is intimately related to theories of persuasion, social influence, nudging and attitude change [30, 8, 34]. From an AI point of view, cognitive planning can be seen as a generalization of epistemic planning [2, 27]. It is not merely a belief state that the planning agent tries to induce but, more generally, a cognitive state or

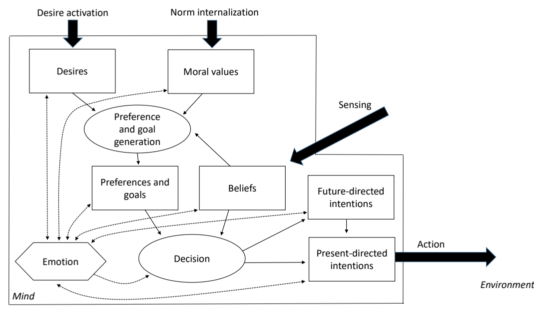

Fig. 1. Cognitive architecture [20]

a behavior of the target agent. The goal of the planning agent is not necessarily communicative in the Gricean sense [15]. Moreover, the kind of communication involved in cognitive planning could be purely behavioral [4] without the use of a codified (verbal or non-verbal) language. For example, to dissuade Bob from disturbing her, Ann can decide to leave her office door ajar. Ann has not necessarily a communicative goal or intention. She may simply want Bob not to knock on her office door. To achieve her goal, she relies on Bob's deductive capabilities. Specifically, she knows that if Bob sees that the door is ajar, he will infer that Ann is busy and, consequently, he will refrain from knocking on the door since he does not want to disturb her when she is busy.

Cognitive planning has also several important ethical implications since building an artificial system with sophisticated persuasive capabilities can be potentially risky (e.g., the system could engage in a manipulative behavior or induce others to do dangerous or illegal actions or persuade them into false beliefs).

In this paper, we situate cognitive planning in the context of a human-machine interaction (HMI) in which the influencer is an artificial communicative agent and the target agent is a human. In particular, our aim is to provide the guidelines for designing, formalizing and then implementing an embodied agent endowed with cognitive planning and normative reasoning capabilities to be used in the context of persuasive multimodal interaction. We will show that in order to develop such agent for HMI applications, it is necessary to integrate logic-based automated reasoning with machine learning techniques in a principled manner at different levels of the design process. The paper is organized as follows. In Section 2, we offer a bird eye view of the overall methodology. Then, we illustrate how cognitive planning can be modeled and automated in an artificial agent using logic-based methods (Section 3), and how to relate cognitive planning to a psychological theory of attitude change in the context of persuasive verbal communication (Section 4). Then, in Section 5, we move from verbal to non-verbal communication: we explain how to use machine learning methods to identify the degree of persuasiveness of a non-verbal message and to endow the agent with persuasive non-verbal capabilities. Finally, Section 6 is devoted to the problem of evaluating the persuasiveness of a virtual agent. In Section 7, we conclude by illustrating some challenges for future research.

2 Methodological foundation

Figure 2 schematically presents a general methodology for designing cognitive planning systems for persuasive multimodal interaction. The integration of ver-

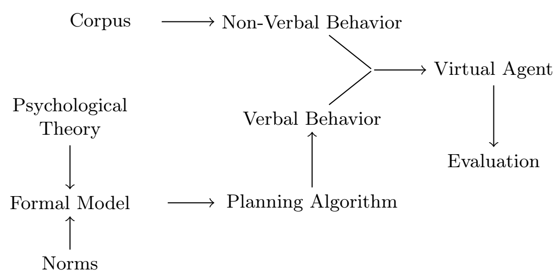

Fig. 2. Overview of the methodology

bal and non-verbal communication is a key feature of the methodology. Indeed, several research studies, particularly in the human-human interaction field, have explored the efficiency of behavioral cues to increase persuasiveness (e.g. [3]). In the domain of virtual agents, empirical research has shown the importance of combining non-verbal cues with verbal ones to create a persuasive virtual agent [13, 7]. On the verbal side, the agent should be able to generate and put into place dialogue plans depending on the context of interaction. Such functionality of the artificial agent can be be achieved satisfactorily by leveraging on logic-based reasoning and planning methods. On the non-verbal side, it is crucial to identify the multimodal cues that should be coupled with the agent's dialogue moves to increase the persuasiveness of its message. This necessarily requires an empirical analysis.

There is an important difference between the two levels. While the agent's generation of dialogue plans relies on psychological theories (top-down approach), the specification of the non-verbal behavior is grounded on an empirical corpus-based analysis (bottom-up approach). This difference is justified by two aspects. On the one hand, there is no well-established theory of non-verbal persuasive communication which clearly identifies the behavioral signals (types, duration, frequencies, combinations) increasing the persuasiveness of the message. Thus, we can only exploit inductive methods, namely, apply machine learning methods to an existing corpus of non-verbal communication in order to predict the persuasiveness of non-verbal signals. On the other hand, we have well-established psychological theories and methods validated by experts for attitude change and persuasion based on verbal communication. Thus, we can rely on them to attain a fine-grained control of the verbal output by the cognitive planning agent. In particular, for the planning agent to be able to generate effective persuasion or influence plans, it should have both i) information about the target agent's overall cognitive state as well as ii) a theory of the causal relationships between the target agent's epistemic and motivational attitudes, emotions and behavior. The latter is usually called Theory of Mind (ToM) [14] and should be grounded on a well-established psychological theory. A crucial step of the design process is to develop a formal language with the right expressiveness to be able to formalize the cognitive state of the target agent as well as the principles of the psychological theory. Such language will be then used to specify the theory-grounded planning algorithm of the artificial agent, namely, the generative component of its verbal communication. The formal model will also be used to formalize the normative and ethical requirements with which the agent is expected to comply, during its interaction with the human. This is to guarantee that it will behave ethically and will refrain from performing obnoxious actions. The norm-aligned verbal behavior generated by the planning module should be appropriately coupled and synchronized with the agent's non-verbal behavior. A user perceptive study should be conducted at the end of the process to evaluate the effectiveness and persuasiveness of the agent's multimodal communication.

In the rest of the paper, we will illustrate the methodology in more detail partially drawing on the experience of a project we have worked on over the last few years. The name of the project is CoPains (Cognitive Planning in Persuasive Multimodal Communication)[4]. The work carried out in the project covers only some aspects of the methodology sketched in Figure 2 and of the integration between logic and machine learning which, as underlined in the introduction, we deem fundamental. For instance, the normative and ethical aspects of cognitive planning and persuasive communication were not explored in detail in the CoPains project. Therefore, we will not focus on them in the rest of the article.

3 Formal model of cognitive planning

In the CoPains project we generalized epistemic planning, in which only epistemic attitudes (beliefs and knowledge) of agents are modeled, to cognitive planning in which the global cognitive state of the target agent is taken into consideration and influenced by the planning agent. A logical language for cognitive planning was developed [10]. The language builds on previous work in epistemic logic in which multi-relational Kripke models are replaced by knowledge bases, in order to facilitate the modeling and implementation of artificial rational agents endowed with cognitive attitudes and with a theory of mind [21, 25, 23, 22, 24]. The main advantage of knowledge bases over Kripke models is their computational groundedness and succinctness which make them well-suited for implementation in multi-agent applications. We explored a variety of algorithmic solutions for implementing cognitive planning in an artificial conversational agent including a SAT-based approach and a QBF-based approach. We combined

[4] https://www.irit.fr/CoPains/

the logic-based cognitive planning module of the conversational agent with its belief revision module. This step was key to come up with a conversational agent that is able to deal with the two fundamental aspects of dialogue, namely, to compute a communicative action or plan and to revise its beliefs in the light of the interlocutor's response.

Two applications were explored during the CoPains project: the first one in the domain of sport recommendation, and a second one in the domain of human-machine collaborative gaming. In the first application, the artificial agent has to motivate the human to practice a physical activity regularly and to help her to find the sport that best suits her preferences and needs. In the second application, the agent has to play a collaborative card game with the human in which reasoning about the human's beliefs and performing informative actions aimed at changing her beliefs play a crucial role. Here, we focus on the first application whose details can be found in [26]. In the next section we will show how the psychological theory of attitude change was imported into the formal model of cognitive planning to endow the agent with an in-depth and theoretically grounded knowledge of how to motivate the human to engage in a regular physical activity. More details about the second application can be found in [19, 11].

4 Formalization of the psychological theory

There exists numerous psychological theories in the literature that could serve as a basis for designing the persuasive verbal behavior of an artificial agent whose ultimate goal is to induce attitude or behavior change in its human interlocutor. In the CoPains project we relied on Motivational Interviewing (MI) [28], a counseling method aimed at empowering the individual by raising her ability to recognize her internal motivators, make reflected decisions and monitor their progress. MI is based on motivation theories to convert the willingness to change into intention and action [16]. The MI communication process leads the person to explore her values, feelings and beliefs about the risky behavior so as to resolve any ambivalence that may arise between desires and inhibitors of change. Throughout the conversation, the MI practitioner infuses a collaborative spirit (partnership) with a non-judgmental and accurate empathetic stance (acceptance). Resolving conflicting motivations through evocation is a central component of MI. Thus, the interviewer's attitude should reflect compassion, respect the interlocutor's opinions and state of readiness to change, avoiding the temptation to give a premature advice. It assumes that people progressively go through various states of mind throughout the process. At the beginning of a session, an individual may express more thoughts in favor of the old habit than the desired one (called sustain talk). The success of the intervention resides in the practitioner's ability to highlight that the weight of the benefits of change is greater than the status quo using the client's own arguments.

Most psychological theories of attitude change and persuasion, including MI, make use of "mentalistic" concepts such as belief, desire, preference and intention, so-called cognitive attitudes, and explain human behavior in the light of these cognitive attitudes. For this reason, the language and algorithm of cognitive planning described in Section 3, in which cognitive attitudes are explicitly represented, have proven to be particularly suitable for formalizing the MI theory and making it operational. They were used to make the artificial planning agent capable of i) ascribing and reasoning about the cognitive attitudes of the human, and ii) predicting the effects of its informative actions on the human's cognitive state and behavior, in accordance with the psychological theory. For example, in light of the MI theory, the agent knows that the first step in motivating the human to practice regular physical activity is to make her aware of the fact that her sedentary lifestyle is in conflict with her goal of being healthy.

As emphasized in Section 2, the psychological theory and its encoding in the formal model of cognitive planning only handle the verbal aspect of persuasive communication. To handle the non-verbal aspect, we mainly relied on machine learning methods which is what we report in the next section.

5 Persuasive non-verbal behavior

Developing a persuasive behavior model implies several challenges, and in particular the precise identification of the behavioral cues associated to persuasion. Indeed, even if the literature in psychology highlights certain cues, the existing research is not sufficient to create a computational model of non-verbal persuasive behavior. One method to identify how a virtual agent should behave to be perceived as persuasive consists in exploring the persuasiveness of human communication in audiovisual corpora. Machine learning techniques are particularly relevant for this task, provided that a corpus with annotations of perceived persuasiveness is available. The POM corpus [29] is, to the best of our knowledge, the only multimedia corpus with annotations of perceived persuasiveness. The POM corpus contains web videos of individuals discussing diverse topics in front of a camera. In a machine learning approach, particular attention must be paid to the interpretability of the model to be able to identify features that can be easily understood and controlled on virtual agents. Different classifiers could be considered such as the traditional SVMs or Random Forests particularly suitable to handle high-dimensional data with a high generalization power [12, 31]. In the CoPains project, we have explored different sets of features and classifiers on the POM corpus to identify the important behavioral cues of persuasion including facial expressions (AUs), head movements and vocal cues [6].

Once the behavioral cues of persuasion have been identified, the next step is to construct the computational model to automatically generate the behavior of a persuasive virtual character. In the CoPains project, on the basis of the persuasive behavioral cues identified from the POM corpus, we have proposed a dictionary to establish reference points that reflect persuasive non-verbal behavior. A convolution-based model, based on this dictionary, was then developed and integrated in a tool to compute the persuasive behavior of the virtual agent based on a video of a neutral face. The tool that we have developed, called THRUST (from neuTral Human face to peRsUaSive virTual face), automatically generates the head movements and facial expressions of a persuasive virtual character. The tool is designed to automatically convert a human's video into a video of a virtual character that exhibits a persuasive non-verbal behavior. Specifically, the tool extracts automatically the human's head movements and facial expressions, applies modifications based on a proposed computational model, and reproduces the resulting head and facial movements on a virtual face. In the next section, we deal with the issue of evaluating the persuasiveness of a virtual agent.

6 Evaluation of virtual persuasiveness

The persuasiveness of a virtual character's behavior must be evaluated through subjective and objective measures. The subjective evaluation refers to the perception of the agent's persuasiveness and believability whereas objective evaluation aims at assessing the change in the user's attitude (see, e.g., [33]). The importance of the subjective evaluation must not be underestimated. It consists in measuring to which degree the behavior of the agent is perceived by users as a persuasive behavior. The most popular method for this consists in asking participants to watch pre-recorded videos of the agent interacting with a user and to indicate their perception through direct questions on the virtual speaker's persuasiveness (e.g. "Did you find the character in the video persuasive?"). The advantage of such videos is twofold. First, it is easier to build an online experimentation and access enough participants for statistical evaluation of the results (which is always a difficulty in subjective studies). Secondly, we can verify that the videos are correctly classified as persuasive or non-persuasive by the classifier described in Section 5. However, the real challenge is to conduct the evaluation in a situation in which the user interacts actively with the virtual agent and not just passively by watching videos. This can be difficult to achieve in practice.

Note that in all cases, a control condition must also be built, in which the agent does not include any persuasive component: the verbal behavior simply follows a predefined list of questions and the non-verbal behavior uses either no transformation or randomly generated non-verbal cues. The subjective evaluation then validates the proposed approach showing that the videos or interaction sessions are perceived as significantly more persuasive than the ones obtained without the model. Testing each modality separately is also interesting to understand what is most important in the user's perception. However, it increases the number of conditions, and by consequence the number of required participants (the standard is to have a minimum of 30 participants per condition). It is also recommended that we evaluate the virtual agents' behavior considering different human videos as input, both female and male, to show that the model provides persuasive output whatever human is in the input video and whatever agent gender in output. Videos generated with virtual agents with different appearances must be evaluated to completely assess the efficiency of the model.

The objective evaluation also raises several challenges by itself. Two different things can be objectively measured: the behaviour change intention (declarative/subjective measure) after the intervention, and the actual behavior change in the long term with objective measures (*e.g.* number of steps per day, time of physical activity per week, *etc.*). In the CoPains project, we measured physical activity through the Global Physical Activity Questionnaire (GPAQ) endorsed by the World Health Organization [5]. This measure of the activity planning gives some view of the intention to change. We also measured the self-efficacy feeling [17], i.e., the individual's belief in his or her capacity to produce specific attainments (in our case, having a regular physical activity).

The real challenge however, we did not address in the CoPains project, is to conduct ecological studies, i.e., having people use the agent outside of the laboratory. Such evaluation situation is prone to system errors (bugs or misconception), irregular use of the system by the other, or other external disruptions. In such ecological conditions, the interaction with the persuasive agent can have a significant positive effect on barrier self-efficacy, but only a limited one on activity planning [18]. This remains however the ultimate goal of persuasive agents.

7 Challenge

We conclude by briefly discussing a crucial challenge for future research: the integration of cognitive planning and Large Language Models (LLMs).

Conversational agents based on LLMs exhibit extraordinary capacities to understand the human interlocutor's utterances and to interact with her in a meaningful and informative way. Nonetheless, they have no representation of the human's cognitive state and have no control on the effects of their actions on it, e.g., on the emotions and stress they may induce in the human. This is what the cognitive planning approach provides: a top-down goal-driven approach to communication in which what the agent should do/not do in a given situation depends on its representation of the human interlocutor's cognitive state and on its theory of the interlocutor's mind. So, a crucial challenge is to combine logic-based cognitive planning models with LLMs for natural language processing (NLP) to be able to exploit the potentialities of both approaches. This challenge is particularly timely, as recent LLM-based agents [35] offer some promising perspectives of integration between logic-based models and natural language communication. The integration can occur at both levels, the reception and the transmission level. For example, it is in principle possible to request to the LLM to extract a formal representation of the human's natural language utterance which can then be processed by the agent's logic-based reasoning module. The planning module can then compute the high-level informative act in response to the user's message which can in turn be transformed, using a second request to the LLM, into a contextualized natural language utterance. Additionally, sentiment analysis models [9] could be exploited to analyze the emotional content of the generated utterance (e.g., to verify the absence of "stressful" markers and expressions).

References

1. Albarracín, D., Sunderrajan, A., Lohmann, S., M. Sally Chan, D.J.: The psychology of attitudes, motivation, and persuasion. In: Albarracín, D., Johnson, B.T. (eds.) The Handbook of Attitudes, Volume 1: Basic Principles, 2nd Edition, pp. 1–42. Routledge (2018)
2. Bolander, T., Andersen, M.B.: Epistemic planning for single- and multi-agent systems. Journal of Applied Non-Classical Logics **21**(1), 9–34 (2011)
3. Burgoon, J.K., Birk, T., Pfau, M.: Nonverbal behaviors, persuasion, and credibility. Human communication research **17**(1), 140–169 (1990)
4. Castelfranchi, C., Pezzulo, G., Tummolini, L.: Behavioral implicit communication (BIC): communicating with smart environments. International Journal of Ambient Computing and Intelligence (IJACI) **2**(1), 1–12 (2010)
5. Chauvin, R., Clavel, C., Sabouret, N., Ravenet, B.: A virtual coach with more or less empathy: impact on older adults' engagement to exercise. In: Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents. pp. 1–9 (2023)
6. Cherni, A., Bertrand, R., Ochs, M.: From neutral human face to persuasive virtual face, a new automatic tool to generate a persuasive attitude. In: Advances in Signal Processing and Artificial Intelligence (ASPAI 2022) (2022)
7. Chidambaram, V., Chiang, Y.H., Mutlu, B.: Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In: Proceedings of the seventh annual ACM/IEEE international conference on Human-Robot Interaction. pp. 293–300 (2012)
8. Cialdini, R.B.: Influence: science and practice. Allyn & Bacon (2001)
9. Clavel, C., Callejas, Z.: Sentiment analysis: from opinion mining to human-agent interaction. IEEE Transactions on affective computing **7**(1), 74–93 (2015)
10. Fernandez Davila, J., Longin, D., Lorini, E., Maris, F.: A simple framework for cognitive planning. In: Proceedings of the Thirty-Fifth AAAI Conference on Artificial Intelligence (AAAI 2021). pp. 6331–6339. AAAI Press (2021)
11. Fernandez Davila, J., Longin, D., Lorini, E., Maris, F.: A logical modeling of the yōkai board game. AI Communications, The European Journal on Artificial Intelligence (forthcoming)
12. Forman, G., Cohen, I.: Learning from little: Comparison of classifiers given little training. In: European Conference on Principles of Data Mining and Knowledge Discovery. pp. 161–172. Springer (2004)
13. Ghazali, A.S., Ham, J., Barakova, E.I., Markopoulos, P.: Poker face influence: persuasive robot with minimal social cues triggers less psychological reactance. In: 2018 27th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN). pp. 940–946. IEEE (2018)
14. Goldman, A.I.: Simulating Minds: The Philosophy, Psychology, and Neuroscience of Mindreading. Oxford University Press (2006)
15. Grice, H.P.: Studies in the way of words (3rd edition). Harvard University Press (1989)
16. Hardcastle, S.J., Hancox, J., Hattar, A., Maxwell-Smith, C., Thøgersen-Ntoumani, C., Hagger, M.S.: Motivating the unmotivated: how can health behavior be changed in those unwilling to change? (2015)
17. Kim, J., Eys, M., Robertson-Wilson, J.: 'if they do it, so can i': a test of a moderated serial mediation model of descriptive norms, self-efficacy, and perceived similarity for predicting physical activity. Psychology & Health **36**(6), 701–718 (2021)

18. Koring, M., Richert, J., Lippke, S., Parschau, L., Reuter, T., Schwarzer, R.: Synergistic effects of planning and self-efficacy on physical activity. Health Education & Behavior **39**(2), 152–158 (2012)
19. Longin, D., Lorini, E., Maris, F.: Beliefs, time and space: A language for the yōkai board game. In: PRIMA 2020: Principles and Practice of Multi-Agent Systems - 23rd International Conference, Nagoya, Japan, November 18-20, 2020, Proceedings. Lecture Notes in Computer Science, vol. 12568, pp. 386–393. Springer (2020)
20. Lorini, E.: Logics for games, emotions and institutions. IfCoLog Journal of Logics and their Applications **4**(9), 3075–3113 (2017)
21. Lorini, E.: In praise of belief bases: Doing epistemic logic without possible worlds. In: Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18). pp. 1915–1922. AAAI Press (2018)
22. Lorini, E.: Exploiting belief bases for building rich epistemic structures. In: Moss, L.S. (ed.) Proceedings Seventeenth Conference on Theoretical Aspects of Rationality and Knowledge (TARK 2019). EPTCS, vol. 297, pp. 332–353 (2019)
23. Lorini, E.: Rethinking epistemic logic with belief bases. Artificial Intelligence **282** (2020)
24. Lorini, E., Rapion, E.: Logical theories of collective attitudes and the belief base perspective. In: Proceedings of the 21st International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2022). ACM (2022)
25. Lorini, E., Romero, F.: Decision procedures for epistemic logic exploiting belief bases. In: Proceedings of the 18th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2019). pp. 944–952. IFAAMAS (2019)
26. Lorini, E., Sabouret, N., Ravenet, B., Fernandez Davila, J., Clavel, C.: Cognitive planning in motivational interviewing. In: Proceedings of the 14th International Conference on Agents and Artificial Intelligence (ICAART 2022). pp. 508–517. SCITEPRESS (2022)
27. Löwe, B., Pacuit, E., Witzel, A.: DEL planning and some tractable cases. In: Proceedings of the 3rd International International Workshop on Logic, Rationality and Interaction (LORI 2011). pp. 179–192. Springer Berlin Heidelberg (2011)
28. Miller, W.R., Rollnick, S.: Motivational interviewing: Helping people change. Guilford press (2012)
29. Park, S., Shim, H.S., Chatterjee, M., Sagae, K., Morency, L.P.: Computational analysis of persuasiveness in social multimedia: A novel dataset and multimodal prediction approach. In: Proceedings of the 16th International Conference on Multimodal Interaction. pp. 50–57 (2014)
30. Rashotte, L.: Social influence. In: Ritzer, G., Ryan, J.M. (eds.) Concise Blackwell Encyclopedia of Sociology. Blackwell (2009)
31. Salperwyck, C., Lemaire, V., de Bois, D.U.d.P.: Impact de la taille de l'ensemble d'apprentissage: une étude empirique. Confrence Internationale Francophone sur l'Extraction et la Gestion de Connaissance (2011)
32. Searle, J.: Rationality in Action. MIT Press, Cambridge (2001)
33. Siegel, M.S.: Persuasive robotics: how robots change our minds. Ph.D. thesis, Massachusetts Institute of Technology (2008)
34. Thaler, R.H., Sunstein, C.R.: Nudge: Improving decisions about health, wealth, and happiness. Yale University Press (2008)
35. Xi, Z., Chen, W., Guo, X., He, W., Ding, Y., Hong, B., Zhang, M., Wang, J., Jin, S., Zhou, E., et al.: The rise and potential of large language model based agents: A survey. arXiv preprint arXiv:2309.07864 (2023)