



HAL
open science

Robotics, Artificial Intelligence and Humans: A Roadmap, or A Cheat Sheet (or Both?)

Aurélie Clodic, Anke Brock, Hélène Cochet, Ophélie Carreras, Raphaëlle Roy

► To cite this version:

Aurélie Clodic, Anke Brock, Hélène Cochet, Ophélie Carreras, Raphaëlle Roy. Robotics, Artificial Intelligence and Humans: A Roadmap, or A Cheat Sheet (or Both?). Robophilosophy Conference 2024, Aug 2024, Aarhus, Denmark. hal-04669136

HAL Id: hal-04669136

<https://enac.hal.science/hal-04669136v1>

Submitted on 15 Nov 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Robotics, Artificial Intelligence and Humans: A Roadmap, or A Cheat Sheet (or both?)

Aurélie CLODIC ^{a,1}, Anke M. BROCK ^b, Hélène COCHET ^c, Ophélie CARRERAS ^c
and Raphaëlle N. ROY ^b

^aLAAS-CNRS, Université de Toulouse, CNRS, Toulouse, France

^bFédération ENAC ISAE-SUPAERO ONERA, Université de Toulouse, France

^cCLLE, UMR5263, Toulouse University, CNRS, UT2J, France

Abstract. Artificial Intelligence and Robotics are trending domains these days. Week after week, there are new impressive videos of robots who are walking, running or flipping, and AI agents performing complex tasks. Such achievements, in addition to all that is happening in related areas, could make us think that the introduction of robots into our daily lives may happen, if not tomorrow, in the near future. But from our robotics research laboratory experience, this future is not for tomorrow. But who actually knows?

In any case, regardless of what will happen, it is important to prepare and be prepared for it, it is important to pursue interdisciplinary fundamental research, it is important to place the Human at the heart of our research concerns. In this paper, we propose a Fundamental and Holistic approach of Artificial Intelligence and Robotics FOR, AMONG, WITH and BY Humans.

Keywords. artificial intelligence, robotics, social robotics, human-robot interaction

1. Motivation

High-level Expert Group on Artificial Intelligence setup by the European Commission proposed that [1]:

Artificial intelligence (AI) systems are software (and possibly also hardware) systems designed by humans that, given a complex goal, act in the physical or digital dimension by perceiving their environment through data acquisition, interpreting the collected structured or unstructured data, reasoning on the knowledge, or processing the information, derived from this data and deciding the best action(s) to take to achieve the given goal. AI systems can either use symbolic rules or learn a numeric

¹Corresponding author: Aurélie Clodic, LAAS CNRS, 7 avenue du Colonel Roche 31031 Toulouse Cedex 4, France; E-mail: aurelie.clodic@laas.fr. This work has been supported by the French National Research Agency (ANR) with AI4HRI (ANR-20-IADJ-0006) and DISCUTER (ANR-21-ASIA-0005) projects.

model, and they can also adapt their behaviour by analysing how the environment is affected by their previous actions.

and then that "Robotics can be defined as 'AI in action in the physical world (also called embodied AI)'."

Artificial Intelligence and Robotics are trending domains these days. Week after week, there are new impressive videos from Boston Dynamics² of walking, running and even flipping robots, while Agility Robotics³ prepares the next generation of robotics automation, and Google presents PALM-E⁴ [2] which takes advantage of Large Language Model to perform complex tasks in the real world. Such achievements, in addition to all that is happening in related areas, could make us think that the introduction of robots into our daily lives may happen, if not tomorrow, in the near future. In the meantime, from our research laboratory experience, this future is not for tomorrow. But who knows?

In any case, regardless of what will happen, it is important to prepare and be prepared for it, it is important to pursue interdisciplinary fundamental research, it is important to place the Human at the heart of our research concerns. At this point, we are convinced that research exchange among experienced researchers from different domains is what is needed and what will bring best results. What is necessary is to allow time to digest each other's research to build upon for the best of Artificial Intelligence and Robotics research FOR, AMONG, WITH and BY Humans.

- Artificial Intelligence and Robotics FOR Humans: Humans should be at the heart of our concerns. How to build trust and acceptance of those systems? How to enable transparency (i.e., being able to understand the why and how of every system's decisions)? How to consider human well-being using those systems (in particular, but not only regarding vulnerable user groups)? How to not hide possible exploitation of humans for dataset generation, as well as preventing unattended personal data collection?
- Artificial Intelligence and Robotics AMONG Humans: Systems are developed for being used in our social world (at home, at work, etc.). How may these systems model and understand these social contexts? What does it mean for a system/agent to be "social" (and does it make sense)? How to take care about bias and cultural differences? How to consider the ethical dimension?
- Artificial Intelligence and Robotics WITH Humans: Systems (at least some of them) are developed to be used during and for interaction with Humans. How could we take advantage of models of human-human interaction such as joint action (not to anthropomorphize, but to better understand human expectations)? How to manage and exploit communication in all dimensions, from sensori-motor to conversational ones? How to deal with knowledge management and consider perspective taking?
- Artificial Intelligence and Robotics BY Humans: Those systems are developed by us, roboticists, which means we are part of the equation. The work of the three first axes will boost the development of our cognitive architecture for robotic systems, but in the meantime, it is also our responsibility to take care of the concerns that may be raised.

²<https://www.bostondynamics.com/atlas>

³<https://agilityrobotics.com/>

⁴<https://palm-e.github.io/>

Our objectives are the following:

**Propose a Fundamental and Holistic approach of Artificial Intelligence and Robotics FOR, AMONG, WITH and BY Humans,
Inform and Participate to the Design and Implementation of the next generation of Robotic Cognitive Architecture FOR, AMONG, WITH and BY Humans.**

Our contribution is related to the conference topic "Prospects for the integration of Humanities and Social Science research into research and development of technology". It proposes some insights regarding why progress in AI and robotics will make expertise from Humanities indispensable, and also where and how the theoretical implications of AI-driven social robots will create new relevance for philosophical (Humanities) expertise.

2. Artificial Intelligence and Robotics FOR Humans

How to design systems that can operate in a responsible way FOR Humans?

In line with other works such as [3,4,5], the European Commission [6] proposes ethical guidelines for trustworthy AI: "Develop, deploy and use AI systems in a way that adheres to the ethical principles of: respect for human autonomy, prevention of harm, fairness and explicability", "Pay particular attention to situations involving more vulnerable groups (...)", "Acknowledge that, while bringing substantial benefits to individuals and society, AI systems also pose certain risks and may have a negative impact (...)". Considering these guidelines in the development of our systems is a challenge that we need to tackle. Of course, several attempts exist.

Already in [7], Astrid Weiss proposed the USUS Evaluation Framework for Human-Robot Interaction where the evaluation concerns were based on usability, social acceptance, user acceptance and societal impact. In [8], Virginia Dignum proposed the ART framework for Accountability, Responsibility and Transparency and proposed the Design for Values principle. In [9,10], Johanna Seibt proposed the Integrative Social Robotics (ISR) framework. This framework defines five principles (*the process principle, the quality principle, the complexity principle, the context principle, and the "values first" principles*) to guide the design and the implementation of social robotics application (but the proposed framework may have broader applications). A complete design-implementation of the framework has been described in [11].

However, while we are aware of these works, we never achieved to operationalize these concepts in our own robotics architecture [12,13,14]. Such objectives are nonetheless becoming more and more explicit worldwide: the European AI Act⁵ acknowledges, for example, that certain AI systems create risks that we must address to avoid undesirable outcomes and proposes a dedicated regulation.

⁵<https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai>

3. Artificial Intelligence and Robotics AMONG Humans

Research Question: How to design systems that can operate AMONG Humans in complex social environments?

Our social world is highly complex—the social context of human-robot interaction consists of the multiplicity of roles, norms, conventions, and social practices that we, as humans, explicitly or implicitly define to handle our daily lives together. Even though roboticists did acknowledge this complexity from the very beginning of social robotics, e.g. [15], we still lack a general theoretical framework for describing a social interaction context.

Researchers from social sciences [16], [17], [18], [19] have offered frameworks for the description of what can be called “asymmetric social interactions” (Seibt) between humans and robots. Brinck et al. [20] propose mutual recognition as a key concept explaining that “Recognition is key to successful cooperation with robots because once mutual recognition has been established between agents: (1) the others become resources for me in a way that promotes equality and sharing, which in turn enables (2) more and other information to be shared and exchanged in HRI”.

In [21], Sciutti et al. proposed to “humanize human-robot interaction” and that robots need to become considerate of humans. They proposed the necessity of bi-directional mutual understanding, which can be viewed as the complement of mutual recognition. They argue that we need, on one side, robots able to understand us and at the same time robots that we can easily understand and anticipate. We have also taken part of this by proposing a set of “basic” requirements for social agency [22], by proposing the use of social practices as a potential model [23] or showing how commitments theory can make sense in a social robotics context [24]. However, these works are still first (design) steps that need to be pursued. In that perspective, besides finding the right way to conceptualize, works must be done to enable robotics platforms to “understand something as something” where “observables are not just detected, they must be read as indicators” [25].

4. Artificial Intelligence and Robotics WITH Humans

Research Question: How to design systems that can operate and more specifically interact WITH Humans through action and communication?

We think that what we call “joint action” in the sphere of human-human interaction may be a source of inspiration for the design of interactive systems. We do not aim to anthropomorphize the system behavior, but to be able to understand how a human approaches and deals with interaction. They are for example several coordination processes at work during a joint action such as self-other distinction, joint attention, shared task representation or intentional action understanding [26]. We have made some propositions about how these processes can make sense in human-robot interaction [27,28,29] and proposed a first implementation [30] in the framework of the JointAction4HRI project⁶.

Although we agree with Curioni et al. [31], when proposing joint action in humans as a model for human-robot interaction, we must admit that operationalizing joint action

⁶<https://jointaction4hri.laas.fr/results>

concepts in an architecture is still an open challenge. We think we need to pursue this initiative by envisioning the overall joint action in all its dimensions and especially among them joint attention, sensori-motor and higher level communication, perspective taking, planned and emergent coordination, language, action execution, adjustment, adaptation, predictability and commitments. As explained in [32]:

currently, while social robotics has dedicated a big effort to equip robots with specific capabilities [...] ever-increasing specialization [...] sometimes prevents researchers from considering the joint action as a whole. [...] For instance, the lack of communicative strategies dedicated to establishing mutual recognition before joint action may trigger previous unrealistic expectations of the human, which can undermine prediction during the whole process. These difficulties can be overcome by adopting an integrated approach of HRI, i.e., an approach that provides us with a way to avoid compartmentalization.

5. Artificial Intelligence and Robotics BY Humans

Research Question: How to design systems that can be effectively implemented BY Humans?

It is crucial to figure out if the fundamental work done along the three other axes can be implemented in a robotics architecture at the end. This means that ideas and concepts that may come up can find their place in a robotics architecture.

In addition, those systems are developed by us, roboticists, which means we are part of the equation. "Robots are not inherently good or bad" [33]. It is also our responsibility to take care of the concerns that may be raised. Robotics engineers must have in mind relevant ethical and societal questions when it comes to the implementation. They can be helped in that process using methods for rapid ethical deliberation [34]. Also, the set of questions proposed by [9] or more recently by [33] could be used to evaluate what could be the benefits but also the damage that the developed system can allow. The participation of stakeholders should be considered from the design phase and the question must be discussed "Should such systems be developed and used in the first place?" [25], but also [9] and [33].

6. Conclusion

In this contribution, we propose to discuss how philosophy, and social sciences at large must accompany roboticists in their work and which dimensions it can take along four axes. Our contribution is related to the conference topic "Prospects for the integration of Humanities and Social Science research into research and development of technology" and proposes some insights regarding why will progress in AI and robotics make Humanities expertise indispensable, and also where and how do the theoretical implications of AI-driven social robots create new relevance for philosophical (Humanities) expertise.

It is important to engage or pursue inter-disciplinary fundamental research, such as the one proposed by Robophilosophy conference [35,36] or by the volume [5]. It is important to place the Human at the heart of our research concerns. We are convinced

that research exchange among experienced researchers is what is needed and what will bring best results. What is necessary is to let us time to digest each other research to build upon for the best of Artificial Intelligence and Robotics research FOR, AMONG, WITH and BY Humans.

References

- [1] A definition of Artificial Intelligence: main capabilities and scientific disciplines. Independent High-Level Expert Group on Artificial Intelligence setup by the European Commission; 2019. Available from: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=56341.
- [2] Driess D, Xia F, Sajjadi MSM, Lynch C, Chowdhery A, Ichter B, et al. PaLM-E: An Embodied Multimodal Language Model. In: arXiv preprint arXiv:2303.03378; 2023. .
- [3] Ethically Aligned Design: A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, Version 2. The IEEE Global Initiative for Ethical Considerations in Artificial Intelligence and Autonomous Systems;. Available from: <https://standards.ieee.org/industry-connections/ec/ead-v1.html>.
- [4] Bryson J, Winfield A. Standardizing Ethical Design for Artificial Intelligence and Autonomous Systems. *Computer*. 2017;50(5):116-9.
- [5] Braunschweig B, Ghallab M. Reflections on Artificial Intelligence for Humanity. Springer International Publishing; 2021. Available from: <http://dx.doi.org/10.1007/978-3-030-69128-8>.
- [6] Ethics guidelines for Trustworthy AI. European Commission and Directorate-General for Communications Networks, Content and Technology; 2019.
- [7] Weiss A, Bernhaupt R, Tscheligi M. The USUS evaluation framework for user-centered HRI. In: *New Frontiers in Human–Robot Interaction*. John Benjamins; 2011. p. 89-110. Available from: <https://www.jbe-platform.com/content/books/9789027283399-ais.2.07wei>.
- [8] Dignum V, Dignum F, Vázquez-Salceda J, Clodic A, Gentile M, Mascarenhas S, et al. Design for Values for Social Robot Architectures. In: *Envisioning Robots in Society – Power, Politics, and Public Space*; 2018. p. 43-52. Available from: <https://hal.laas.fr/hal-01943831>.
- [9] Seibt J. "Integrative Social Robotics": A New Method Paradigm to Solve the Description and the Regulation Problem? In: Seibt J, Nørskov M, Schack Andersen S, editors. *What Social Robots Can and Should Do*. *Frontiers in Artificial Intelligence and Applications*. IOS Press; 2016. p. 104-15. Available from: <http://conferences.au.dk/robo-philosophy/>.
- [10] Seibt J, Damholdt MF, Vestergaard C. Five Principles of Integrative Social Robotics. In: *Robophilosophy/TRANSOR*; 2018. .
- [11] Fischer K, Seibt J, Rodogno R, Rasmussen MK, Weiss A, Bodenhausen L, et al. Integrative Social Robotics Hands-on. *Interaction Studies*. 2020.
- [12] Lemaignan S, Warnier M, Sisbot EA, Clodic A, Alami R. Artificial Cognition for Social Human-Robot Interaction: An Implementation. *Artificial Intelligence*. 2017 Jun;247:45-69. Available from: <https://hal.archives-ouvertes.fr/hal-01857498>.
- [13] Chatila R, Renaudo E, Andries M, Chavez-Garcia OR, Luce-Vayrac P, Gottstein R, et al. Toward Self-Aware Robots. *Frontiers in Robotics and AI*. 2018 Aug;5:88. Available from: <https://hal.archives-ouvertes.fr/hal-01856931>.
- [14] Sarthou G, Mayima A, Buisan G, Belhassein K, Clodic A. The Director Task: a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures. In: *30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021)*. Vancouver (Virtual), Canada; 2021. Available from: <https://hal.archives-ouvertes.fr/hal-03327222>.
- [15] Breazeal C. Toward sociable robots. *Robotics and Autonomous Systems*. 2003;42(3):167-175. *Socially Interactive Robots*. Available from: <http://www.sciencedirect.com/science/article/pii/S0921889002003731>.
- [16] Seibt J. Varieties of the "As If": Five Ways to Simulate an Action. *Frontiers in Artificial Intelligence and Applications*. 2014;273(Sociable Robots and the Future of Social Relations):97-104. Available from: <https://doi.org/10.3233/978-1-61499-480-0-97>.
- [17] Seibt J, Vestergaard C, Damholdt MF. In: *Sociomorphing, Not Anthropomorphizing: Towards a Typology of Experienced Sociality*. vol. 335 of *Frontiers in Artificial Intelligence and Applications*. IOS Press; 2020. p. 51-67. Available from: <https://doi.org/10.3233/FAIA200900>.

- [18] Fiebich A, Nguyen N, Schwarzkopf S. In: Misselhorn C, editor. *Cooperation with Robots? A Two-Dimensional Approach*. Cham: Springer International Publishing; 2015. p. 25-43. Available from: https://doi.org/10.1007/978-3-319-15515-9_2.
- [19] Misselhorn C. *Collective agency and cooperation in natural and artificial systems : explanation, implementation and simulation*. Cham: Springer; 2015.
- [20] Brinck I, Balkenius C. *Mutual Recognition in Human-Robot Interaction: A Deflationary Account*. *Philosophy and Technology*. 2018;1(1):53-70.
- [21] Sciutti A, Mara M, Tagliasco V, Sandini G. *Humanizing Human-Robot Interaction: On the Importance of Mutual Understanding*. *IEEE Technology and Society Magazine*. 2018;37(1):22-9.
- [22] Fernández Castro V, Hakli R, Clodic A. *What does it take to be a social agent?* In: Appeared in: Nørskov, M., Seibt J., Quick O. 2020. *Culturally Sustainable Social Robotics-Proceedings of Robophilosophy 2020*. Series *Frontiers of AI and Its Applications*, IOS Press, Amsterdam.; 2020. Available from: <https://hal.archives-ouvertes.fr/hal-02963044>.
- [23] Clodic A, Vázquez-Salceda J, Dignum F, Mascarenhas S, Dignum V, Augello A, et al. *On the Pertinence of Social Practices for Social Robotics*. In: Press I, editor. *Envisioning Robots in Society – Power, Politics, and Public Space*; 2018. p. 36-74. Available from: <https://hal.laas.fr/hal-01943774>.
- [24] Fernández Castro V, Mayima A, Belhassein K, Clodic A. In: Gransche B, Bellon J, Nähr-Wagener S, editors. *The Role of Commitments in Socially Appropriate Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg; 2024. p. 223-48. Available from: https://doi.org/10.1007/978-3-662-68021-6_11.
- [25] Bellon J, Eyssel F, Gransche B, Nähr-Wagener S, Wullenkord R. *Theory and Practice of Sociosensitive and Socioactive Systems*. Springer Fachmedien Wiesbaden; 2022. Available from: <http://dx.doi.org/10.1007/978-3-658-36946-0>.
- [26] Pacherie E. *The Phenomenology of Joint Action: Self-Agency vs. Joint-Agency*. In: Seemann A, editor. *Joint Attention: New Developments*. MIT Press; 2012. p. 343-89. Available from: https://jeannicod.ccsd.cnrs.fr/ijn_00778370.
- [27] Clodic A, Pacherie E, Alami R, Chatila R. *Key Elements for Human Robot Joint Action*. In: *Sociality and Normativity for RobotsPhilosophical Inquiries into Human-Robot Interactions*. *Studies in the Philosophy of Sociality*. Springer; 2017. p. 159-77. Available from: https://hal.laas.fr/ijn_03084126.
- [28] Clodic A, Alami R. *What Is It to Implement a Human-Robot Joint Action?* In: *Robotics, AI, and Humanity*. Springer International Publishing; 2021. p. 229-38. Available from: <https://hal.laas.fr/hal-03184727>.
- [29] Belhassein K, Víctor FC, Mayima A, Clodic A, Pacherie E, Guidetti M, et al. *Addressing joint action challenges in HRI: Insights from psychology and philosophy*. *Acta Psychologica*. 2022 Feb;222:103476. Available from: <https://hal.archives-ouvertes.fr/hal-03505047>.
- [30] Sarthou G, Mayima A, Buisan G, Belhassein K, Clodic A. *The Director Task: a Psychology-Inspired Task to Assess Cognitive and Interactive Robot Architectures*. In: *30th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN 2021)*. Vancouver (Virtual), Canada; 2021. Available from: <https://hal.archives-ouvertes.fr/hal-03327222>.
- [31] Curioni A, Knoblich G, Sebanz N. *Joint Action in Humans: A Model for Human-Robot Interaction*. In: *Humanoid Robotics: A Reference*. Springer Netherlands; 2018. p. 2149-67. Available from: https://doi.org/10.1007/978-94-007-6046-2_126.
- [32] Belhassein K, Fernández Castro V, Mayima A. *A Horizontal Approach to Communication for Human-Robot Joint Action: Towards Situated and Sustainable Robotics*. In: *Culturally Sustainable Social Robotics*. IOS Press; 2020. p. 204-14.
- [33] Šabanović S, Charisi V, Belpaeme T, Bethel CL, Matarić M, Murphy R, et al. *“Robots for good”: Ten defining questions*. *Science Robotics*. 2023;8(84):ead4238. Available from: <https://www.science.org/doi/abs/10.1126/scirobotics.adl4238>.
- [34] Steen M, Neef M, Schaap T. *A Method for Rapid Ethical Deliberation in Research and Innovation Projects*. *International Journal of Technoethics*. 2021 Jul;12(2):72–85. Available from: <http://dx.doi.org/10.4018/IJT.2021070106>.
- [35] Nørskov M, Seibt J, Quick OS. *Culturally Sustainable Social Robotics: Proceedings of robophilosophy 2020*. Amsterdam, NY: IOS Press; 2021.
- [36] Hakli R, Mäkelä P, Seibt J, editors. *Social Robots in Social Institutions*. IOS Press; 2023.