Benefits and Risks of AI Companions

Erik Derné1,2 and Jan Zahálka1

1 CIIRC CTU in Prague, Czech Republic; 2 ELLIS Alicante, Spain

Motivation

• **AI companion** — AI-equipped agent, chatbot, or robot serving as a personal companion to a human through meaningful interaction

• AI companions are emerging in various fields and applications, ranging from **therapeutic to casual**

• Robots are becoming **teammates** rather than tools

• **Feasible** through natural language processing technology (advances in deep neural networks, namely transformers)

• Collaboration in mixed teams of humans and robots brings focus on **social aspects**

• Human in the loop → **adaptive AI, continuous learning**

• Dealing with uncertainty by **asking whenever unsure**

Benefits

• **Always available**, patient, inexhaustible, and tireless; (ideally) reliable

• **Adaptation** and behavior adjustment based on the specific needs of each user

• User model learning and real-time model updating, **emotional analysis** to choose a suitable approach

• Learning from only **few examples** through interaction

• **Positive social impact**: mental health and well-being applications, dealing with loneliness

• **Artificial empathy** in social robotics

• People tend to be more **honest** (open up) to a machine, bypassing the fear of failing and being judged

Risks

• **Misuse** of the personal information coming from the interaction

• Risk of causing **injuries** in case of embodied AI systems

• Adverse effects such as adopting **undesired behavioral patterns**, causing negative emotions, addiction

• **Mental health** implications, high sensitivity of the user in difficult moments of life, dealing with controversial topics

• Risk of **social isolation**, preferring the interaction with AI companions to the interaction with humans

• **Wrong decisions** of the AI system may have severe consequences

Research Questions

1) How does the embodiment and **appearance** of the AI companion influence the acceptance by the user?

2) How to customize the behavior and **adapt** to the needs of each individual user?

3) In which way should the companion’s verbal and nonverbal **communication** be designed to fit the situation?

4) To make themselves trustworthy and transparent, how can AI companions **explain** their reasoning?

5) What are the ways to deal with **uncertainty** of the AI systems?

6) How can AI **support decision-making** without rendering human knowledge and skills obsolete?

7) Which information about the user, the task, and the environment should be sensed and which should be inferred to diminish **privacy** concerns of the user?

8) What are the techniques to foster the **personal development** of humans instead of reducing their involvement and learning to rely blindly on their AI companions?

9) In which way can AI companions overcome **human cognitive biases** in decision making?

10) What are the implications of **pretending to be a human** teammate in contrast to disclosing openly the AI nature of the companion?

11) **Who decides what ethical behavior is?**

12) How to ensure the **security** of the data collected by the AI companion?

13) **Who is responsible** for the decisions of the AI system?

References


Acknowledgment

This work was supported by the European Regional Development Fund under the project Robotics for Industry 4.0 (reg. no. CZ.02.1.01/0.0/0.0/15_003/0000470).

Images credit: RIKEN-TRI Collaboration Center for Human-Interactive Robot Research and WikimediaCommons.