Lara Sakarya
L.Sakarya@student.tudelft.nl

Supervisors: Chirag Raman and Ojas Shirekar

Applications of The Active Inference and The Free-Energy Principle Frameworks for Mimicking Social Human Behaviours on Intelligent Agents

a Systematic Literature Review

1. Introduction

- **Active inference (AIF)** is a theory of human perception, planning, and action
- **Free Energy Principle (FEP)** optimizes surprise
- Minimization of surprise through **variational free energy (VFE)** using generative models
- **Expected free energy (EFE)** depicted in Figure 1 extends VFE
- Collaborative agents, collective intelligence, and modelling of trust in human-computer interactions

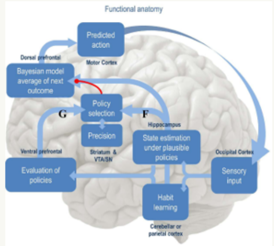

Figure 1: Anatomy of belief updating [1]

2. Research Questions

How have active inference and the free-energy principle been applied to embodied agents and the mimicking of social human behaviours?

RQ 1. What methods and models have been used to apply the active inference and the free-energy principle to embodied agents?
RQ 2. How do the active inference and the free-energy principle relate to social human behaviours?
RQ 3. What are the challenges and limitations of active inference and the free-energy principle applications on intelligent agents?

3. Methods

- Databases
  - Cell Press
  - Scopus
  - IEEE Xplore
  - Web of Science
  - ACM Digital Library
- Phases are depicted in Figure 2

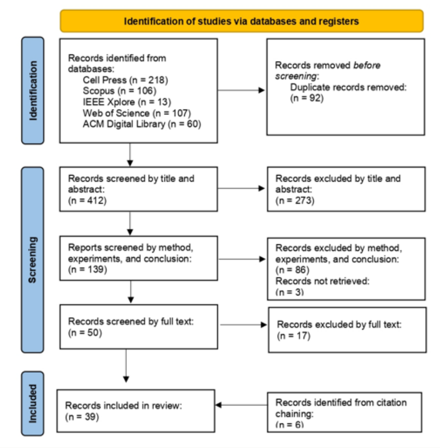

Figure 2: PRISMA Flow Diagram

4. Results – Models and Applications

- Habit formation
- Goal-directed behaviour
  - Probabilistic Programming
  - Bayesian Target Modelling for Active Inference
  - Deep Active Inference (DAI)
- Exploratory and exploitative behaviour
  - Imitation Learning
  - DAI
  - World Models
  - Predictive Coding (PC)
  - Sophisticated AIF

Figure 3: Applications found in literature

- Social interactions
  - Predictive Coding (PC)
  - DAI
- Decision-making
  - A free energy model
  - Boltzmann distribution
  - Chance constraint
- Applications on robots
  - Multimodal Variational Autoencoder Active Inference (MVAE-AIF)
  - Multimodal Hierarchical Dirichlet Process (MHDP)

4. Results – Mimicking Social Human Behaviour

- Collaborative AIF agents for team-optimal behaviour [3]
- Emergence of social rules [4]
- Heterogeneous agent teams [5]
- Collective intelligence [6]
- Trust in human-computer interactions [7]
- Cumulative culture [8]
- Emotion recognition [2]

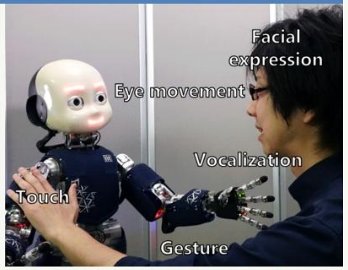

Figure 4: Multimodal human-robot interaction [2]

4. Results – Validity and Limitations

- Translation to complex real-life applications
  - Models built on assumptions
  - Limited social interactions
  - **High model complexity and computational demands**
  - Simple tasks
  - **Exclusion of real-life dynamics**
- Limitations with data
  - **Insufficient training data**
  - **Synthetic data usage**
- Model Validity
  - Small sample size
  - Problems with extension to complex environments
  - Clarity over complexity
  - Construction of the generative model

5. Discussion

- Various **machine learning** methods are combined with AIF to build adaptive agents with **epistemic, social, and collaborative behaviour.**
- Applications of AIF in simple environments indicate that these agents' abilities can be **extended to navigate social contexts.**
- Contex-sensitive behaviour can enable seamless and efficient communication with humans.
- AIF models become more complex when applied to more intricate behaviours and contexts, leading to inapplicability to real-life scenarios. In theory, agents can **acquire knowledge with their intentions** if these challenges are overcome.

6. Conclusion

- AIF and FEP applications can provide insight into human cognition and elevate intelligence and social behaviour in intelligent agents.
- Current AIF models should be applied to real-life scenarios to ensure their applicability.
- Future improvements should focus on integrating physical dynamics and computational aspects into simple AIF models.

7. References

[1] K. Friston, T. FitzGerald, F. Rigoli, P. Schwartenbeck, J. ODoherty, and G. Pezzulo. Active inference and learning. Neuroscience and Biobehavioral Reviews, 68:862–879, 2016.

[2] Takato Horii, Yukie Nagai, and Minoru Asada. Active perception based on energy minimization in multimodal human-robot interaction. In HAI'17, page 103-110, New York, NY, USA, 2017. Association for Computing Machinery. journalAbbreviation: HAI'17.

[3] G. Levchuk, K. Pattipati, D. Serfaty, A. Fouse, and R. McCormack. Active inference in multi-agent systems: context-driven collaboration and decentralized purpose-driven team adaptation. volume 2018-March, pages 157–165, 2018.

[4] M. Hartwig and A. Peters. Cooperation and social rules emerging from the principle of surprise minimization. Frontiers in Psychology, 11, 2021.

[5] G. Levchuk, A. Fouse, K. Pattipati, D. Serfaty, and R. McCormack. Active learning and structure adaptation in teams of heterogeneous agents: Designing organizations of the future. volume 10653, 2018.

[6] R. Kaufmann, P. Gupta, and J. Taylor. An active inference model of collective intelligence. Entropy, 23(7), 2021.

[7] F. Schoeller, M. Miller, R. Salomon, and K.J. Friston. Trust as extended control: Human-machine interactions as active inference. Frontiers in Systems Neuroscience, 15, 2021.

[8] N. Kastel, C. Hesp, K.R. Ridderinkhof, and K.J. Friston. Small steps for mankind: Modeling the emergence of cumulative culture from joint active inference communication. Frontiers in Neurorobotics, 16, 2023.