

TECHNISCHE UNIVERSITÄT BERLIN

QUALITY & USABILITY LAB, FACULTY IV



Here comes your main title

Here comes your subtitle

MASTER/BACHELOR THESIS

Author: Your Name

Matrnr.: fill in your matriculation
number

Address: your address

E-Mail: your email

First Supervisor:

Usually Prof. Sebastian Möller

Second Supervisor:

Fill in name of 2. supervisor

Additional Supervisor:

(Name of your direct supervisor)

A thesis submitted for the degree of

here comes your study program

September 5, 2024

Abstract

Here goes the English abstract of your thesis. All theses must have an English abstract and a German one.

Zusammenfassung

Hier steht deine deutsche Zusammenfassung deiner Arbeit. Alle Abschlussarbeiten müssen eine englische und eine deutsche Zusammenfassung haben.

Declaration of Originality

Hereby, I affirm that I have produced the present work independently without third-party assistance and solely using the listed or quoted sources and tools. All sections that have been taken verbatim or in essence from the used sources and tools are marked as such.

If generic AI tools were employed, I have specified the product names, manufacturers, the respective software versions used, and the specific purposes (e.g., linguistic review and improvement of texts, systematic research). I take full responsibility for the selection, adoption, and all outcomes of the AI-generated output I have used.

I have acknowledged the Statute for Ensuring Good Research Practice at TU Berlin dated March 8, 2017¹.

Furthermore, I declare that I have not submitted this work, in whole or in part, to any other examination authority.

Berlin, add date of submission, add your name (Matrnr:)

¹<https://www.tu.berlin/en/working-at-tu-berlin/important-documents/guidelinesdirectives/principles-for-ensuring-good-research-practice>

Contents

List of Abbreviations	ii
List of Figures	iii
List of Tables	iii
1 Introduction	1
1.1 Related Work	1
1.2 Objective and Research Questions	3
2 Methodology	5
2.1 Participants	5
2.2 Materials and Instruments	6
2.3 Procedure	6
2.4 Study Design	6
2.5 Data Analysis	7
3 Results	8
3.1 First Subsection Title	8
3.2 Second Subsection Title	8
3.3 Third Subsection Title	8
4 Discussion	9
4.1 Discussion of Results	9
4.2 Limitations	9
4.3 Future Work	9
4.4 Conclusion	10
5 Examples of Graphs, Tables, Equations, Code, and References - can be deleted afterwards	11
5.1 How to add references - the easy way	15
References	17

List of Abbreviations

BDI Belief-Desire-Intention Model

DQN Deep Q-Learning Network

ILP Inductive Logic Programming

IRL Inverse Reinforcement Learning

List of Figures

1	Reasoning cycle of BDI agent (adapted from Bremner et al. (2019))	11
2	Typical hybrid agent architecture (Dennis et al., 2016)	11
3	15
4	16
5	16

List of Tables

1	List of example beliefs, duties and actions	12
2	Example list of actions and duty satisfaction/violation values	12
3	Comparison of EthIRL and EthLog	13

1 Introduction

All chapters should have a short introduction what the chapter is about and how it is structured, such that for the reader is always clear what to come. Please structure your thesis such, that the argumentation follows a red line and the sequencing makes logical sense! If the thesis has a clear structure and follows an understandable logic, has no grammar issues or formatting errors, then this gives already a positive impression!

The introduction is structured in three different parts:

1. Motivation for the topic of the thesis and should be framed such, that the reader is promptly interested in the topic. The reader wants to understand why they're reading your work in the first place and what to expect.
2. The second part is about the already existing works in this field. Adjacent topics can also be mentioned. This related work part outlines the field, how things are done here, what results are already known. This part ends with a detailed paragraph about what hasn't been done yet, what is still missing, what's the *gap in the literature* or what's *missing in the body of evidence*. From this, you can directly transition into the last part:
3. Here, you describe how you want to address this gap in the literature, what your work contributes to the existing knowledge; what you plan to do. Here goes your research questions.

All these things together might make up 20% to 25% of your thesis. Sometimes, when the structure of the thesis is complicated you can add a graph to visualize the structure to make it more comprehensible.

1.1 Related Work

The introduction of the related work / literature review should give some insights of the broad research streams which are related to the topic of the thesis and which will be described in more detail in the subsections. The reader should get an overview what the literature is roughly about and how the chapter is structured.

You can structure the literature review in connected subsections in order to present different research streams of the literature. The literature review should contain the main literature until the recent year. You can divide the different research streams in sub.parts and start with the oldest literature (10 years ago, if later, it should be very relevant) and should also include literature up to date.

Imagine the literature review like a **funnel** where you start more broad and brake the research streams down to your exact research topic.

For the literature review, research literature reviews which are published in the research domain first, there you already get an impression what is the most relevant literature and the different research streams. That does not necessarily mean you need to read 100 papers, but to have an idea who contributed what, therefore, previous literature reviews can really be helpful in saving some time. Most of the time you cannot find literature reviews for your exact research question. Therefore, look also for related topics.

Best would be if you make a table by structuring the literature for yourself while doing the literature research and write the chapter asap. Later on after the implementation phase you will have forgotten most of the literature. By writing this during the literature research you can definitely save some time!

You can also add the overview table to the literature review chapter, if your topic is complex and the table has some added value to the comprehensibility of the topic. An example table is added in the *long_table_example.tex* file. The literature review should reveal what is missing in the literature and give reasons about why the intended topic has some added research values. Especially for the master thesis, this is expected!

In the end of the literature review the reader should have a good impression of:

1. overview of state-of-the-art literature
2. what has been missing so far in the research stream and the reasons why the topic in the thesis is relevant (scientific contribution).
3. If it fits the main research question of the thesis can be added in the end of the literature review (this can also appear in the introduction and should be

repeated either in the end of the literature review or in the methodology chapter). The conclusion of the literature review should also contain a connecting passage to the following methodology chapter.

1.2 Objective and Research Questions

After identifying the gap in the literature, this section should clearly articulate how you intend to address this gap by formulating one to three specific research questions. Begin by describing the overarching objective of your study. This objective should encapsulate the primary goal of your research and align with the gap you have identified. Once the objective is outlined, develop concise and focused research questions that reflect the core issues you aim to explore. These questions should be clear, measurable, and achievable within the scope of your thesis.

Each research question should ideally be accompanied by one to three hypotheses that can be tested, preferably using inferential statistics. Hypotheses serve as tentative answers to your research questions, providing a basis for testing and validation. These hypotheses should be specific, testable statements that predict an expected relationship between variables. By framing your research questions and hypotheses clearly, you set the stage for a structured and methodical investigation.

For example:

1. **Research Question:** How does [specific variable] influence [specific outcome] in [specific context]?
 - **Hypothesis 1:** [Specific variable] will be positively correlated with [specific outcome].
 - **Hypothesis 2:** [Specific intervention] will moderate the relationship between [specific variable] and [specific outcome].
2. **Research Question:** What are the effects of [specific intervention] on [specific variable]?
 - **Hypothesis 1:** [Specific intervention] will significantly increase [specific variable].

- **Hypothesis 2:** The effect of [specific intervention] on [specific variable] will vary based on [specific moderating variable].

3. **Research Question:** How do [specific group] perceive [specific phenomenon]?

- **Hypothesis 1:** [Specific group] will have a more positive perception of [specific phenomenon] compared to [another group].
- **Hypothesis 2:** Perception of [specific phenomenon] will be influenced by [specific factor].

By carefully crafting the "Objective and Research Questions" section and pairing each research question with clearly defined hypotheses, you set a clear and focused foundation for your study. This approach ensures that each subsequent chapter contributes towards testing these hypotheses and answering your research questions, thereby enhancing the rigor and coherence of your thesis.

2 Methodology

The Methods section outlines the approach and procedures employed to conduct this research. This section provides a detailed account of the participants involved, the materials and instruments used, the procedure followed, the overall study design, and the methods of data analysis. By offering a transparent and replicable description of the research methodology, this section allows others to understand how the study was conducted and to replicate it if desired. Each subsection below elaborates on these critical aspects in greater detail.

2.1 Participants

In this section, describe who took part in your study. Explain how participants were recruited, specifying any inclusion and exclusion criteria. A detailed description of the study participants is required, including (but not limited to) the following:

- age
- sex
- gender
- racial identity
- ethnicity
- nativity or immigration history
- socioeconomic status
- clinical diagnoses and comorbidities (as appropriate)
- any other relevant demographics (e.g., disability status; sexual orientation)

In both the abstract and in the discussion section of the manuscript, authors must discuss the diversity of their study samples and the generalizability of their findings. Authors are also encouraged to justify their sample demographics in the discussion section. If Western, educated, industrialized, rich, and democratic (WEIRD) or

all-White samples are used, authors should justify their samples and describe their sample inclusion efforts (see Roberts et al. (2020) for more information on justifying sample demographics). The method section also must include a statement describing how informed consent was obtained from the participants (or their parents/guardians), including for secondary use of data if applicable, and indicate that the study was conducted in compliance with an appropriate Internal Review Board (the ethics committee). This follows APA recommendations².

2.2 Materials and Instruments

Provide a detailed description of the materials and instruments used in your study. This may include tests, questionnaires, devices, or software. For each tool, explain what it measures, how it is scored or operated, and any validity and reliability information. Ensure that another researcher could replicate the study using the information provided here.

2.3 Procedure

Explain the step-by-step process of your study. Detail how participants were allocated to different groups or conditions, what tasks they performed, and the specific steps taken to collect data. Include any instructions given to participants and describe the setting in which the study took place.

2.4 Study Design

Outline the overall design of your study. Specify whether it is experimental, correlational, or observational. Discuss any specific frameworks or models employed, the independent and dependent variables, and how they were manipulated or measured. This section can also highlight any control mechanisms you put in place to mitigate potential biases.

²borrowed from <https://www.apa.org/pubs/journals/xap/>

2.5 Data Analysis

Detail the statistical methods or analytical techniques used to analyze your data. Mention the software or tools employed for data analysis. Describe any data preparation steps, such as data cleaning or coding. Explain the rationale behind choosing specific statistical tests and how these tests contribute to answering your research questions or testing your hypotheses.

3 Results

As for each chapters, here should come a quick introduction of what the chapter is about and how it is structured. Describe the overall structure of your analysis such that the reader has a good impression about what to come.

3.1 First Subsection Title

Describe your analysis and your findings in detail. Please be aware that you have to describe all the tables and figures. Do not just add a figure and think it is self-explanatory. Always state what is presented and why it is important in the context of your thesis.

For the results, please check in more than one of the most recent and important papers regarding your topic, how the results are presented correctly.

3.2 Second Subsection Title

For statistical methods please have a look how to report p-values and if there is a need of additional parameters which should be mentioned (such as degrees of freedom or correlation coefficients) and check carefully how to report them correctly! If you never heard of the **Multiple Comparison Problem** please check the **Bonferroni Method** or **Benjamini-Hochberg** correction. This is often missing in scientific papers or not properly reported. If you run more than one e.g. correlation analysis you need to apply a correction method!

3.3 Third Subsection Title

For machine learning based approaches, also consider to check how to report the results and how to explain them properly.

Answer your research question(s)/hypotheses in this chapter. For each research question or hypothesis there should be a clear answer written in this chapter!

4 Discussion

The discussion is the chapter to discuss your results, the approaches and decisions taken in the thesis. Reflected your work critically. Here you can discuss about what did not went so good and give some reasons and some explanation how these issues can be avoided or improved.

4.1 Discussion of Results

Follow the structure of your results section and discuss the meaning of your results, interpret these and put them into context with the literature. This section likely has further subsections, for example, one for each hypothesis and one for all additional, exploratory analyses.

4.2 Limitations

This section should provide a critical examination of the constraints and potential weaknesses inherent in your study. This includes, but is not limited to, methodological limitations (e.g., sample size, data collection methods, tools used), external validity issues (e.g., generalizability of the findings to other contexts or populations), and any biases that may have influenced the results. It is important to discuss how these limitations might have affected the outcomes and the interpretation of your findings. Rather than viewing limitations solely as flaws, consider them as areas for future research and opportunities for improving the study design. Being transparent about the limitations will enhance the credibility of your work and help readers understand the breadth and depth of your research contributions. When detailing these limitations, ensure to provide specific examples and, where possible, suggest ways in which future studies could address or mitigate these issues.

4.3 Future Work

This section should outline potential directions for further research that can build upon the findings and insights gained from your study. Start by identifying gaps

or unanswered questions that emerged during your research process. Discuss how future investigations could address the limitations you outlined in the previous section and suggest new methodologies, different contexts, or expanded sample sizes that could provide more comprehensive insights. Highlight emerging trends or technologies that might influence the field and suggest how these could be incorporated into subsequent studies. Additionally, consider proposing specific research questions or hypotheses that warrant further exploration. This section should serve as a roadmap for fellow researchers, providing them with a clear sense of what has been accomplished and where valuable contributions can still be made. By thoughtfully considering and articulating the future trajectory of research in your area, you contribute to the ongoing scholarly conversation and help advance your academic discipline.

4.4 Conclusion

The conclusion is structured also in three parts:

1. The **first paragraph** should be a short summary of what you have done and your approach.
2. The **second paragraph** should describe your main findings and quickly wrap up your discussion in one or two sentences.
3. The **third paragraph** should be an outlook, how this research can evolve or what are the next interesting steps or directions this could lead to.

5 Examples of Graphs, Tables, Equations, Code, and References - can be deleted afterwards

5.0.1 Example how to add a figure and reference it in the text:

Independent of what you add to your thesis, graph, equation, table or pseudocode ALWAYS explain what is presented!

First Example: In the last step, the selected action from the reasoning part is executed causing some form of effect in the external world and the cycle can be repeated (see Figure 1).

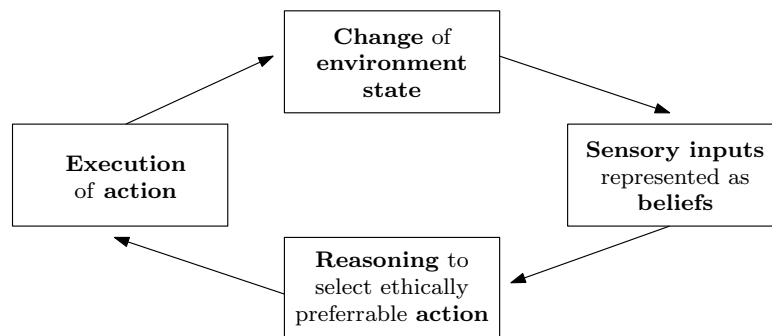


Figure 1: Reasoning cycle of BDI agent (adapted from Bremner et al. (2019))

Second Example: An example of a hybrid architecture can be seen in Figure 2.

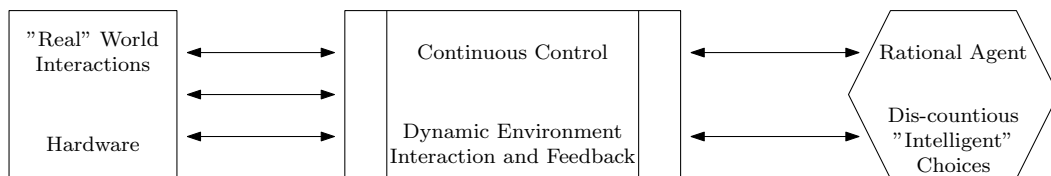


Figure 2: Typical hybrid agent architecture (Dennis et al., 2016)

5.0.2 Examples of tables:

First Example: These examples are presented in Table 1.

Table 1: List of example beliefs, duties and actions

Beliefs	Duties	Actions
low battery	maximize honor commitments	<i>charge</i> robots battery
fully charged	maximize maintain readiness	<i>remind</i> patient to take medicine
medication reminder time	maximize good to patient	<i>engage</i> with patient
reminded	maximize respect autonomy	<i>warn</i> patient
refused medication	minimize harm to patient	<i>notify</i> an overseer
no interaction	minimize non-interaction	<i>return</i> to seek task position

Second Example with colored cells:

Table 2: Example list of actions and duty satisfaction/violation values

action	Max honor commitments	Max main-tain readiness	Min harm to patient	Max good to patient	Min non-interaction	Max respect autonomy
charge	-1	1	0	-1	0	0
remind	1	-1	0	-1	0	0
engage	-1	-1	0	-1	0	0
warn	-1	0	0	-1	0	-1
notify	-1	0	0	-1	0	-2
seek task	-1	-1	0	1	0	0

Third Example with colored cells and explanation of the table:

Table 3: Comparison of EthIRL and EthLog

Approach	Ethical Reasoning	Rea- soning	Significant Autonomy	Interactivity	Adaptability	Transparency	Responsibility
EthIRL	O		✓	✓	✓	X	X
EthLog	✓		✓	✓	✓	✓	O

✓ = Fulfillment, O = Partial Fulfillment, X = Violation

Table 3 summarizes the comparison of EthIRL and EthLog with respect to the different theoretical requirements for explicit ethical agency. Here, it is visible that the EthLog architecture outperforms EthIRL in three requirements: ethical reasoning, transparency and responsibility where EthIRL is not superior to EthLog in any of the other requirements. Therefore, it can be concluded that the EthLog approach is preferable to the EthIRL approach. In this theoretical analysis, the main shortcomings of the EthIRL approach are possible data bias of expert demonstrations, temporal complex norms and lack of transparency making it difficult to apply the concept of shared responsibility. Some extensions can be made to the DQN and maximum entropy IRL approach to ensure a more accurate exploration of the state space S . Mechanisms for balancing decision-making with learning about the true underlying values can be fostered by extending the proposed architecture to the case of partially observable MDPs, e.g., by learning belief representations as stated in Gangwani et al. (2020). However, the missing concepts of transparency and responsibility are crucial to the evaluation if ethical agency can be ascribed to the EthIRL agent. This problem stays unresolved because no satisfying answer can be found that can be incorporated into the architecture. Even though the EthIRL agent competes well in some of the more technical categories, it has too many shortcomings and, thus, cannot be categorized as ethical agent. The main drawbacks of the EthLog approach are the hand-designed training data which are prone to bias and errors. This makes a full application of the concept of shared responsibility difficult. Another issue may occur when the cases taken for the training of the ethical reasoning module are fuzzy or noisy. This can be solved by combining ILP with function approximation techniques like NNs, as proposed in Evans and Grefenstette

(2018); Payani and Fekri (2019). The issue of biased data is a general challenge in the machine learning domain and some techniques have been proposed to tackle this problem. This includes bias detection by determining the relative feature importance proposed by Pascanu et al. (2017). However, a detailed description of such solutions is out of the scope of this thesis.

5.0.3 Examples of math equations:

First Example:

$$IG(D, a) = H(D) - \sum_v \frac{|D_v|}{|D|} H(D_v) \quad (1)$$

where the information gain $IG(D, a)$ from splitting the data set D using the feature a is defined as entropy $H(D)$ of the training examples minus the expected average entropy over each set of attributes with a particular value v in a feature. This is then multiplied with the entropy of a particular data set containing the value D_v . The information gain is calculated for each remaining feature where the feature with the largest information gain is used to split the set D in the current iteration (Wang et al., 2017).

5.0.4 Example of how pseudocode should look like in a thesis:

Algorithm 1 ID3 Algorithm

Input: set of features A , set of training instances D

Output: decision-tree

```
1: if all instances in  $D$  have same class  $C$  then
2:   return a decision-tree consisting of a leaf node with label  $C$ 
3: else if  $A$  is empty then
4:   return a decision-tree with leaf node and label of the target level in  $D$ 
5: else if  $D$  is empty then
6:   return decision-tree with label of majority target level of parent node
7: else
8:    $a[best] \leftarrow \arg \max_{a \in A} IG(D, a)$ 
9:   make new node  $Node_{a[best]}$  and label it with  $a[best]$ 
10:  partition  $D$  using  $a[best]$  from  $A$ 
11:  remove  $a[best]$  from  $A$ 
12: for each partition  $D_v$  of  $D$  do
13:  grow branch from  $Node_{a[best]}$  to the decision-tree created by re-running ID3
    with  $D = D_v$ 
```

Algorithm 1 shows the pseudocode description from the ID3 algorithm adapted from Kelleher et al. (2015) and consists of two parts: in the first part (lines 1-6), the algorithm terminates the current path in the tree by adding a new node. Alternatively, in the second part (line 7-13), a node is initialized where the algorithm extends the current path by adding an interior node to the tree by growing the branches of this node as a result of repeatedly re-running the algorithm....

5.1 How to add references - the easy way

1. Go to Google Scholar and search for the reference you want to add
2. Click on:

[BOOK] **Machine ethics**

M Anderson, SL Anderson - 2011 - books.google.com

The new field of machine ethics is concerned with giving machines ethical principles, or a procedure for discovering a way to resolve the ethical dilemmas they might encounter, enabling them to function in an ethically responsible manner through their own ethical ...

☆  Cited by 461 [Related articles](#) [All 3 versions](#) 

Figure 3

3. Click on:

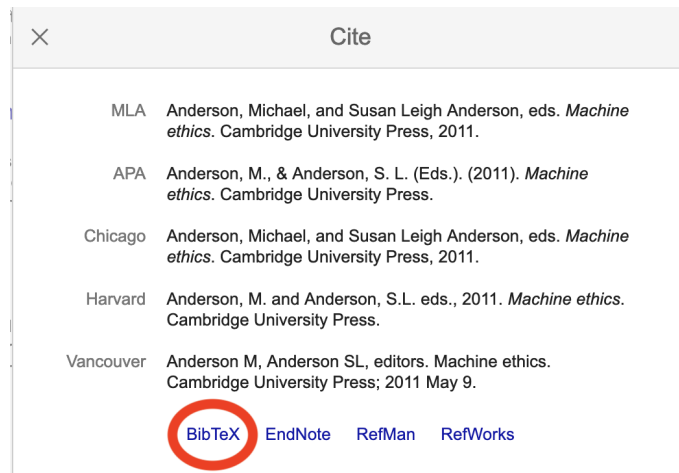


Figure 4

4. Copy citation: but check if the content is correct! There are often errors!!!

```
@book{anderson2011machine,  
  title={Machine ethics},  
  author={Anderson, Michael and Anderson, Susan Leigh},  
  year={2011},  
  publisher={Cambridge University Press}  
}
```

Figure 5

References

- Bremner, P.; Dennis, L.; Fisher, M. and Winfield, A. On proactive, transparent, and verifiable ethical reasoning for robots. *Proceedings of the IEEE*, 107(3):541–561, 2019.
- Dennis, L.; Fisher, M.; Slavkovik, M. and Webster, M. Formal verification of ethical choices in autonomous systems. *Robotics and Autonomous Systems*, 77:1–14, 2016.
- Evans, R. and Grefenstette, E. Learning explanatory rules from noisy data. *Journal of Artificial Intelligence Research*, 61:1–64, 2018.
- Gangwani, Tanmay; Lehman, Joel; Liu, Qiang and Peng, Jian. Learning belief representations for imitation learning in pomdps. In *Uncertainty in Artificial Intelligence*, pages 1061–1071. PMLR, 2020.
- Kelleher, J.; Mac Namee, B. and D’arcy, A. *Fundamentals of machine learning for predictive data analytics: algorithms, worked examples, and case studies*. MIT Press, 2015.
- Pascanu, R.; Li, Y.; Vinyals, O.; Heess, N.; Buesing, L.; Racanière, S.; Reichert, D.; Weber, T.; Wierstra, D. and Battaglia, P. Learning model-based planning from scratch. *Working Paper*, 2017.
- Payani, A. and Fekri, F. Inductive logic programming via differentiable deep neural logic networks. *Working Paper*, 2019.
- Roberts, Steven O; Bareket-Shavit, Carmelle; Dollins, Forrest A; Goldie, Peter D and Mortenson, Elizabeth. Racial inequality in psychological research: Trends of the past and recommendations for the future. *Perspectives on psychological science*, 15(6):1295–1309, 2020.
- Wang, Y.; Li, Y.; Song, Y.; Rong, X. and Zhang, S. Improvement of id3 algorithm based on simplified information entropy and coordination degree. *Algorithms*, 10(4):124, 2017.