

Towards Responsible AI Code Development: A Six-Value Process Model for Junior and Novice AI Programmers

Rick van Kersbergen¹ and Saskia Robben¹

¹Amsterdam University of Applied Sciences, 1091 GH Amsterdam, The Netherlands
`r.van.kersbergen@hva.nl`, `s.m.b.robben@hva.nl`

1 Introduction

The rapid progress of Artificial Intelligence (AI) has brought attention to ethical worries in AI development. AI becomes more common and it's critical that it's designed and used responsibly. This research aims to help new AI programmers understand how to make responsible choices when programming AI by presenting a novel Responsible AI Process Model containing six Responsible AI values: Fairness, Sustainability, Human Centricism, Efficiency, Transparency and Explainability. Also, we developed a Visual Studio Code (VSC) extension that incorporates this process model and assists the programmer cultivate Responsible AI code.

2 Related work

The examination of the Responsible AI (RAI) literature revealed a pronounced overlap of values, albeit with instances where the same value was named differently or provided with varying definitions, indicating a pressing need for consistency in defining and labelling Responsible AI values.

The effectiveness of process models in improving code quality and teaching coding is a grey area. [1] [2] describe steps that need to be completed while developing AI. The steps from both sources come down to defining the problem, preparing the data, building and training the model and evaluating the model. Also, the prominence of RAI values within process models and patterns appears to be relatively limited, which contradicts the importance of values as stated within the literature.

[3] and [4] suggest that using design pattern documentation in the form of pattern comment lines (PCLs) within an IDE can be beneficial for programmers. The conclusion of the research was that available plug-ins primarily focus on general coding, offering auto-completion features that help programmers write code faster but do not necessarily teach responsible coding. Currently, there appears to be an absence of an extension that offers this.

3 Methodology

We conducted qualitative interviews with five experts in the field of RAI. The objective was to identify the key concepts and considerations they deemed critical within the context of RAI. The insights were then compared with the current literature on RAI, which lead to the development of the process model.

To translate abstract values into specific norms, the process Design for Values, proposed by Friedman et al. [5], was utilized. Based on Steen's theory on the differences between intrinsic and instrumental values [6], the aforementioned values could be assigned varying degrees of importance.

For shaping the process model, the final step of the Design for Values approach was utilized [5] by posing value-related questions found within the literature to uphold ethical principles. These questions could be linked to specific instrumental values [7] [8] [9] and translated into tasks that must be executed to address them. Finally, tasks and associated questions related to instrumental values could be assigned to specific steps within the programming process.

An experiment was conducted, involving students from the Bachelor of the Artificial Intelligence program at the University of Amsterdam and Master students from the University of Applied Sciences. The participants were divided a control group and an experiment group. The control group tackled an assignment without the aid of the VSC extension, while the experiment group utilized the extension and followed the process model. Each participant was also interviewed and asked to provide explanations for why they made their choices within the assignment.

4 Results

The methods yielded values with interpretations. Moral imagination, seen by some experts as integral to early project stages involving ethical considerations, is perceived more as a task than a standalone value. Despite this perspective, the experts' insights on moral imagination will be integrated into the process model.

The process model encompasses four programming steps: Define Problem, Construct and Prepare Data, Build and Train Model, and Evaluate Model. Each step identifies the values that should be upheld, and corresponding questions are stated, which simultaneously correlate with the tasks to be performed within each programming step. This process model had then been used in the development of the Visual Studio Code extension, which was utilized in the evaluation of the model.

The results of the evaluation provide evidence of the positive impact of our approach. Participants who used the VSC extension and process model demonstrated a better grasp of Responsible AI values and their integration into programming. Notably, the experiment group exhibited a higher ability to identify sensitive features in data, recognize proxies, consider fairness and bias, choose appropriate algorithms, and comprehend the significance of false negative rates.

5 Conclusion

These findings have important implications for AI education and responsible development. The positive results show that adding Responsible AI values to the development process helps programmers make informed and ethical choices. Our approach combines ethical thinking with technical tasks, helping programmers understand AI development as a whole and allowing them to include values in their code. The outcomes of this research contribute to the body of knowledge by providing further support for the positive impact of process models and patterns on code development capabilities. Additionally, they validate the importance of incorporating Responsible AI values into process models, aligning with existing theories and enhancing the practical application of these values.

References

1. S. Studer, T. B. Bui, C. Drescher, A. Hanuschkin, L. Winkler, S. Peters, and K.-R. Mueller, “Towards CRISP-ML(q): A machine learning process model with quality assurance methodology.”
2. “Responsible AI toolkit.”
3. L. Prechelt, B. Unger, and M. Philippsen, “Two controlled experiments assessing the usefulness of design pattern documentation in program.”
4. M. Vokáč, W. Tichy, D. I. K. Sjøberg, E. Arisholm, and M. Aldrin, “A controlled experiment comparing the maintainability of programs designed with and without design patterns—a replication in a real programming environment,” vol. 9, no. 3, pp. 149–195.
5. B. Friedman and P. Kahn, *Value Sensitive Design Shaping Technology with Moral Imagination*. The MIT Press.
6. M. Steen, *Ethics for People Who Work in Tech*. Crc Press, 1 ed.
7. N. Piersma, *System Error, Please Restart. Hoe we vera twoorde IT-systemen kunnen bouwen*. Eburon.
8. V. Dignum, *Responsible Artificial Intelligence How to Develop and Use AI in a Responsible Way*. Springer, 1 ed.
9. R. Benjamins, A. Barbado, and D. Sierra, “Responsible AI by design in practice,”