# A Use Case Library for Machine Ethics

January 4, 2026

## Research Question

How can examples of ethical reasoning be organised into a taxonomy?

## Background

A number of systems are available that claim to enable a system to reason ethically [3, 4]. Most of these systems have been evaluated on one or two simple ethical problems. This makes it very difficult to compare systems since the sample problems vary wildly from whether it is ethical to give a girl flowers, to whether you should bake someone a cake or kill them, through to the popular trolley problems around collisions in autonomous cars. Compounding this problem is the fact that it is not necessarily clear what the ”correct” answer is in many cases. An initial attempt at creating a benchmark set of ethical problems for machines was reported in [1], but this presented the problems in a largely unstructured fashion and is no longer readily available.

Meanwhile the Trustworthy Autonomous Systems programme created a Use Case library for examples of trust-based reasoning and interaction in AI research [2]. This structures examples according to a number of features such as the kind of interaction involved, the type of trust being considered and so on. Can something similar be created for ethical reasoning examples?

## Approach

The object of this project would be to create a standardised template for describing ethical problems that machines can reason about, implement a database and front end (ideally web-based) for storing and accessing the problems, and encode a representative sample of such problems from the literature in the template and store them in the database. As an extension the problems could then be extracted into some existing machine ethics implementations and comparisons drawn.

## Milestones

1. A template for describing ethical decision problems for use by machine ethics systems.
2. A database system for storing such problems with a representative set of examples within it.
3. Evaluation of existing machine ethics systems on the problem set.

## References

[1] Bjørgen, E. P., Madsen, S., Bjørknes, T. S., Heimsæter, F. V., Håvik, R., Linderud, M., Longberg, P., Dennis, L. A., and Slavkovik, M. Cake, death, and trolleys: Dilemmas as benchmarks of ethical decision-making. In *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society, AIES 2018, New Orleans, LA, USA, February 02-03, 2018* (2018), J. Furman, G. E. Marchant, H. Price, and F. Rossi, Eds., ACM, pp. 23–29.

[2] Masters, P., Young, V., Chamberlain, A., Weerawardhana, S. S., McKenna, P. E., Lu, Y., Dowthwaite, L., Luff, P., and Moreau, L. A practical taxonomy of TAS-related usecase scenarios. In *Proceedings of the First International Symposium on Trustworthy Autonomous Systems, TAS 2023, Edinburgh, United Kingdom, July 11-12, 2023* (2023), ACM, pp. 41:1–41:6.

[3] Tolmeijer, S., Kneer, M., Sarasua, C., Christen, M., and Bernstein, A. Implementations in machine ethics: A survey. *ACM Comput. Surv. 53*, 6 (2021), 132:1–132:38.

[4] Vishwanath, A., Dennis, L. A., and Slavkovik, M. Reinforcement learning and machine ethics: a systematic review. *CoRR abs/2407.02425* (2024).