

DEPARTMENT: EDUCATION & TRAINING

Teaching Ethics as a Creative Subject: Ideas From an Interdisciplinary Workshop

Luke Haliburton , *Ludwig-Maximilians Universität, 80539, Munich, Germany*

Alexander Heimerl, *University of Augsburg, 86159, Augsburg, Germany*

Stephanie Böhme, *Friedrich-Alexander-Universität Erlangen-Nürnberg, 91054, Erlangen, Germany*

Elisabeth André, *University of Augsburg, 86159, Augsburg, Germany*

Albrecht Schmidt , *Ludwig-Maximilians Universität, 80539, Munich, Germany*

Designers and developers have great power and they are often not aware of it. The decisions they make in system design and the way they implement algorithms inevitably encode ethical values. We propose using “Dark Scenarios” as a teaching tool to make ethical issues and value sensitive design a creative process that fits the mindset of designers and developers.

VALUES ARE EMBEDDED IN YOUR DESIGN

Designing and implementing technical systems is inseparably linked to making ethical decisions—but which functionalities are prioritized? Which users do we have in mind when designing the product? What is the underlying business model and how does it affect different user groups? These questions are at the center of the process of creating new products and services.

Typically, developers are pragmatic in dealing with this, they rarely see design as an ethical challenge or value dilemma. In many cases there is a “natural” solution, a clear way of implementation that is not questioned, and alternatives are not explicitly considered. This lack of reflection is pragmatic and should not be confused with a lack of interest. Theoretical consideration of abstract values and ethical concepts is simply not part of the routine for most developers.

We therefore argue that teaching ethics—which has an impact on the actual work developers will do—

needs to take their practice into account and needs to align with how they create products. We created an interdisciplinary workshop to facilitate learning about ethics in an engaging and creative way.

Our experience is that students in computer science, design, and engineering like to create things and enjoy coming up with solutions to tricky, often technical, problem. Similarly, students in psychology and medical fields like to study people and their behaviors to inspire clever and impactful interventions.

When it comes to teaching ethics to these students, they typically do not mind reading and discussing philosophical texts and reasoning about ethical concepts. However, we argue this is so far detached from their subjects that it may have no impact on their professional work as a psychologist or software developer. Furthermore, students should be encouraged to reflect on the ethical implications of technologies at a very early stage of the design process instead of applying repair mechanisms afterward.¹ To enhance the students’ engagement, various methods to ethics education have been proposed, such as holding a trial,² analyzing case studies,³ or employing science fiction.⁴ In our approach, we aim to make learning about ethics a creative task that requires problem solving skills, which plays to their strengths.

There are three major learning goals to our approach. We want students to understand:

- ▶ The power of ethics in creating products and interventions.
- ▶ That value choices are not automatic, they are linked to decisions made by designers.
- ▶ That ethical decision making is not abstract, but is incorporated into each design step and algorithm.

To understand the impact of ethical choices and embedded values, it is crucial to understand the perspective of the end user.

EMPATHY FOR THE USER

Being empathic with the user is a key to creating good solutions. But people are complex! Especially when it comes to value choices, potential users will often not be able to articulate their expectations.

Take this simple example: Assume we create a service where users can either pay by subscription or by providing personal information. If you ask users if they want the product to be free, they will probably say “yes.” If you ask whether the product should collect personal information and keep a profile about them, they will probably say “no.” But clearly these answers do not help the designer make decisions because users choose daily to use common “free” services that collect personal information to generate revenues. People make complex value choices and they often behave differently from what they communicate in an interview (see the psychological phenomena of *social desirability bias*⁵).

Inconsistencies between user statements and behaviors cannot be solved even with great user centered design. The problem is not a lack of understanding of the users’ desires, the problem is that designers and developers must make design tradeoffs, which inherently encode values. They embed their view of what is “right” (and also what makes economic sense and is implementable) in the product.

These embedded values are inherent to the design and development process, so we propose that educating developers on ethics is the key to improving the ethical impact of technologies.

EDUCATE DEVELOPERS, THEY MAKE THE CHOICES

Developers and designers are ultimately responsible for new technologies and how they impact individuals and society. As designers make what may appear to be minor design decisions, they make ethical decisions as well, and thus they need to understand ethics, values, morals, and legal systems. Society expects that they will not be harmed by new products, so the decision makers must ensure that this is true. Value decisions are rarely easy; hence, as educators, we need to give guidance and educate developers.

In our view, the key challenge is to empower developers to analyze systems in a way that enables them to see potentially problematic value choices. Developers need to be able to spot tradeoffs; they need to

identify potential negative impacts both on individuals and on society.

Currently, legal requirements (e.g., GDPR) are regularly checked as a requirement, but ethical the principles behind the laws are often not integrated in the design process. We aim for extended design and development processes that assess potential solutions with regard to encoded values and their ethical implications.

The take home messages for developers and designers are following:

- ▶ Every design decision is a value choice that impacts people! Especially as systems gain more autonomy (e.g., autonomous systems, intelligent systems) these decisions become more far reaching, and it does not matter if they are made explicitly (e.g., creating the algorithm) or implicitly (e.g., by picking the data the system is trained on).
- ▶ There is no recipe for value choices. There is often no “right” ethical decision. Designers and developers will not find the solution on Stack Overflow. Values change with the task and with resources, and ethical considerations are always contextual.

Dark Scenarios are a powerful instrument to help designers and developers become more aware of ethical issues and (implicit) value choices.

DARK SCENARIOS FOR ETHICS AS TOOL

“Dark Scenarios” are powerful framing tools for creatively exploring ethical problems. The technique involves pretending that you are an “evil” person and using that perspective to consider how you could misuse technological designs.

Imagining that you are evil offers a novel and innovative (and sometimes funny) opportunity to discuss a difficult topic. This thought experiment promotes creativity and cooperation, and it sensitizes participants to possible problematic ethical issues. The creative process of looking at things from an evil perspective also helps designers understand the positive aspects of a design. Overall, considering Dark Scenarios can be a powerful tool to better understand potential design solutions. Dark scenarios may also inspire student discussions on the ethical implications of dark patterns in UX⁶ that nudge users into unfavorable behaviors.

Dark Scenarios, in the context of ethics, can be used at any point in the design, development, and deployment process. In an early stage, scenario

descriptions or paper prototypes can be used as input, while in later stages full prototypes are most appropriate. Dark Scenarios can also be used either on a high level, basically from the perspective of the user, or on a deep technical level.

One approach to explore a Dark Scenario is to take a current design and analyze it for weak points in three steps:

1. Identify a question, a dataset, concept, solution, prototype, or product you want to assess and decide on the way to communicate it to the participants. Examples include
 - a. A concept description and screen mockups of a new navigation app.
 - b. A description of a dataset on student performance with example data rows.
 - c. The source code of a prioritization algorithm for a vaccination.
2. Ask participants to pretend they are evil and think about how they could exploit such a system. Here we recommend separating different concerns into different short sessions. For example, one aspect could be the data and technology side (how are data collection, processing, and storage achieved?); an alternative might be on the people involved (what could developers, researchers, participants, administrators, users, etc. do?). For example:
 - a. Using a navigation app to guide people into taking a route that comes by a storefront to drive sales, and ensuring that certain people, such as frequent customers, have a free route at the cost of delays for others.
 - b. Providing help and tutoring to students that have good prospects, while not supporting students that are likely to fail anyway.
 - c. Getting a higher vaccination priority by gaming the scheduling algorithm by studying people who received higher priority to understand the characteristics used by the algorithm.
3. Identify aspects in the design that allow someone to behave in an evil way, as well as who has access to these aspects. For example:
 - a. Software developers.
 - b. Data base administrators.
 - c. Users.

In both approaches, the intention is for participants to see their systems from a new perspective and uncover potential ethical problems. Using creative approaches and engaging methods like

considering Dark Scenarios may be effective ways for designers and developers to learn practical lessons about ethics. To better explore and evaluate our proposed approach, we implemented our suggestions in a one-day multidisciplinary workshop.

INITIAL EXPERIENCE: AN INTERDISCIPLINARY WORKSHOP

We implemented our creative ethics approach in a workshop on design ethics with Ph.D. researchers and professors from multiple fields (e.g., computer science, psychology, and communication). The attendees ($N = 25$) were all part of the ForDigitHealth research group, which represents five German Universities. The workshop was conducted virtually on Zoom due to the coronavirus pandemic.

Prior to the workshop, we sent a survey to the attendees asking about their current practices relating to ethics, security, and privacy. We presented a short summary of this survey as an opening at the beginning of the workshop to provide context about practices within the group.

Following the initial presentation, the attendees participated in a Dark Scenario exercise.

The participants were divided into breakout rooms and given two rounds of prompts. The prompts encouraged participants to think about a particular situation from an evil perspective and imagine what steps they could take to misuse some technology. As an example, the first prompt was:

Imagine: You have a wealth of data from your participants and you are really evil!

What are concrete and actionable things you could do to harm your participants? Is there material to blackmail them? Is there information about them they surely do not want other people to have? Is there information they may be embarrassed about? How could you use this information to make them feel bad? Who could you share this information with to make it even worse?

The next phase was a keynote talk presented by Jessica Heesen, an external expert on ethics, followed by an engaging Q&A session.

Based on our preworkshop survey, we identified data collection, storage, and processing as key research functions where our participants encountered ethical concerns. The next phase of the workshop was therefore focused on each of these steps. We split the attendees into small groups, which were preselected to ensure a range of disciplines in each. We provided a mural board with prompts for each section, such as: "What should you do if you accidentally

learn information about a user that could be harmful to themselves or other people?" The groups discussed the prompts for 20 minutes and then reported back to the main group for one minute after each session.

We sent attendees a postworkshop survey to evaluate the workshop approach. Results ($N = 6$) revealed that the keynote was the most informative session, followed by the Dark Scenarios, and preworkshop survey results. Participants appreciated the external input and also benefited from discussion sessions to reflect on the input and share their own experiences.

CONCLUSION

We suggest using a creative approach to teach ethics in a more impactful and practical way. We successfully applied this approach in an interdisciplinary workshop. Our pilot workshop suggests that promoting creative thought leads to greater understanding and deeper insights. By encouraging participants to think like an "evil" person, they were able to think outside of the box and overcome inhibitions. Intentionally encouraging interdisciplinarity created opportunities for different practices to be shared across fields.

In all, we found that teaching ethics as a creative subject is an effective way to promote discussions and initiate practical learning on what can be a difficult subject.

REFERENCE

1. A. G. Pillai *et al.*, "Co-designing resources for ethics education in HCI," in *Proc. Conf. Hum. Factors Comput. Syst. Extended Abstr.*, 2021, pp. 109:1–109:5, doi: [10.1145/3411763.3441349](https://doi.org/10.1145/3411763.3441349).
2. C. Beaton, "Creative ways to teach ethics and assess learning," in *Proc. 39th IEEE Int. Conf. Frontiers Educ. Conf.*, 2009, pp. 897–900, doi: [10.1109/FIE.2009.5350654](https://doi.org/10.1109/FIE.2009.5350654).
3. E. Burton, J. Goldsmith, S. Koenig, B. Kuipers, N. Mattei, and T. Walsh, "Ethical considerations in artificial intelligence courses," *AI Mag.*, vol. 38, no. 2, pp. 22–34, 2017, doi: [10.1609/aimag.v38i2.2731](https://doi.org/10.1609/aimag.v38i2.2731).
4. E. Burton, J. Goldsmith, and N. Mattei, "How to teach computer ethics through science fiction," *Commun. ACM*, vol. 61, no. 8, pp. 54–64, 2018, doi: [10.1145/3154485](https://doi.org/10.1145/3154485).
5. P. Grimm, *Social Desirability Bias*. Hoboken, NJ, USA: Wiley International Encyclopedia of Marketing, 2010, doi: [10.1002/9781444316568.wiem02057](https://doi.org/10.1002/9781444316568.wiem02057).
6. K. Lukoff, A. Hiniker, C. M. Gray, A. Mathur, and S. S. Chivukula, "What can CHI do about dark patterns?," in *Proc. Conf. Hum. Factors Comput. Syst. Extended Abstr.*, 2021, pp. 102:1–102:6, doi: [10.1145/3411763.3441360](https://doi.org/10.1145/3411763.3441360).

LUKE HALIBURTON is currently working toward the Ph.D. degree in the Media Informatics Group, Ludwig-Maximilians Universität, Munich, Germany. Contact him at luke.haliburton@ifi.lmu.de.

ALEXANDER HEIMERL is currently working toward the Ph.D. degree in the Lab for Human-Centered AI, University of Augsburg, Augsburg, Germany. Contact him at alexander.heimerl@informatik.uni-augsburg.de.

STEPHANIE BÖHME is a postdoctoral researcher at the Department of Clinical Psychology and Psychotherapy, Institute of Psychology, Friedrich-Alexander-Universität, Erlangen-Nürnberg, Germany. Contact her at stephanie.boehme@fau.de.

ELISABETH ANDRÉ is a full professor and the founding chair of Human-Centered Artificial Intelligence, University of Augsburg, Germany. Contact her at elisabeth.andre@uni-a.de.

ALBRECHT SCHMIDT is a professor with the Media Informatics Group, Ludwig-Maximilians Universität, Munich, Germany. Contact him at albrecht.schmidt@ifi.lmu.de.