07.09.2023

CO-BOLD

A Serious Game to Promote a Responsible Use of Artificial Intelligence in Business

Von Johannes Katsarov, Prof. Dr. Paul Drews und Prof.in Dr. Hannah Trittin-Ulbrich

In this blog article, we present the serious moral game "CO-BOLD", which we have developed and tested in the past two years. The goal of CO-BOLD is to sensitize learners to the ethical opportunities and risks of artificial intelligence (AI) in business. First, we discuss the need to promote students' awareness of ethical risks associated with AI. Before this background, we introduce the goals of our innovative game-based training module and provide a brief overview, how the game works. Finally, we present first results from our use of the game in teaching students of Management and Information Science at Leuphana.

1 The Need for AI Ethics Education

Many – if not most – of the recent revolutions of artificial intelligence are based on machine learning. They include the recommendation systems that power services like Netflix, spam-detection software, the automatic generation of artwork based on verbal commands through systems like Midjourney, powerful chatbots like ChatGPT and many others. Machine learning feeds computers with tons of data (e.g., music and information about music) and then tasks the computers with making sense of this data (identifying patterns) through statistical processes. The algorithms that emerge through this process can then be used to fulfil complex tasks, e.g., to create new music, to recognize cancer cells, or to identify people based on their body movements.

Machine Learning bears many benefits for society. Hundreds of millions of people already use relevant applications regularly. However, machine learning also comes with many ethical risks and problems [1-3].

A couple of examples will help to illustrate some of these issues:

1. In the Dutch Childcare Benefits Scandal, an artificial intelligence (AI) run by the Dutch government wrongly accused approximately 26.000 parents of fraud against the Dutch welfare system. Without a fair trial, many of these parents were required to repay sums larger than 10.000 Euros within two weeks – driving people who already depended on social welfare deeper into poverty and ruining the lives of countless families.

2. In Germany – and many other countries – financial algorithms based on machine learning tend to demand higher interest rates and collateral from women than men (if women even get offers). Based on statistical data from the past, the computers assume that women are less creditworthy per se, and even make them worse offers when their incomes etc. are equal to those of men.

3. The newest generation of machine-learning-based chatbots (known as Large Language Models) have become so convincing that they can manipulate and deceive people. For example, a young father recently made a deal with a chatbot, sacrificing his own life in exchange for the AI's "promise" to stop human-made climate change.

AI is not a "fad" that will go over soon. It is an emerging technology that will continue to be used in millions of projects worldwide in all sectors of work to automatize, improve, or revolutionize work processes. At the same time, the named problems show that professionals (e.g., government officials, financial managers, or information scientists) often fail to recognize the ethical risks and problems associated with AI. This observation is seconded by our research. At Leuphana, we developed an ethical sensitivity test for risks associated with artificial intelligence. In this test, people read a short dialogue about an innovative fitness application, and then list the ethical issues that they noticed in the text. Before training, a group of 28 Master students of management only scored an average of 2.8 points (from 14) on this test. In other words, from seven ethical issues that could be recognized, most respondents noticed fewer than two problems, and were oblivious to the rest (including, for instance, the potential of severe harm). This lack of ethical sensitivity is common (we have also found it in students of information science) and a serious problem. After all, today's students of Management and Information Science will be tomorrow's decision makers. Among others, they will decide what technologies their organizations will adopt and develop in the future.

2 What are the Challenges for Ethics Education? How can we address them?

Traditional ethics education has two major shortcomings. First, ethics education usually focuses on the development of ethical reasoning, judgment, and problem-solving [4]. Without doubt, moral reasoning is an important competence, an ability that helps people in sorting out dilemmatic situations. However, this ability only accounts for 10% of failures in moral behavior [5]. We need to also train people to…

1. Care enough to look closely (due diligence, moral commitment),

2. Recognize ethical problems intuitively (moral sensitivity), and to

3. Speak up and address problems pro-actively (moral resoluteness) [6].

Secondly, ethics education is most frequently based on lectures and reading ("theoretical approach") or the discussion of cases ("deliberative approach") [4]. However, these approaches are significantly less effective than experiential learning ("engagement approach"), which ask learners to imagine how they would actually deal with challenging situations and provides them with feedback on their imagined behaviors. A recent meta-analysis on research-ethics training found that an engagement approach led to better learning outcomes of all sorts – including knowledge acquisition, moral reasoning, attitudes, ethical sensitivity, and moral behavior [10].

Figure 1: Screenshot from the final level of CO-BOLD

3 Design of the Serious Game CO-BOLD

To provide effective ethics training on the responsible use of artificial intelligence at Leuphana, we have developed a serious moral game named "CO-BOLD". CO-BOLD has been designed to address the main reasons, why people often fail to detect and address ethical problems in the usage of AI. In playing the game, learners learn to recognize eight ethical risks and problems that are common in applications [1-3]. Moreover, they learn to think about the quality of AI applications in terms of eight quality dimensions [7] and to develop a vantage of professional skepticism [8] in evaluating AI applications. Finally, the game challenges learners to speak up and address ethical problems with AI against business pressures [9].

To achieve these learning goals, CO-BOLD has players slip into the role of a quality assurance manager at a large tech company. Their mission is to assess the quality of an innovative AI assistant for investment consultants ("CO-BOLD") before it is delivered to a large bank. To win the game, players need to adopt a questioning mind. They must find out, what kinds of problems CO-BOLD to prevent a tragic ending. And they need to persuade others to uphold quality standards against business interests.

The game experience is structured into four levels.

1. The first level introduces players to their role as a quality manager at a big tech company called "e-matur technologies". The get to know their team, which is composed of an information scientist and a business analyst who will conduct investigations on players' command. Additionally, they get to know the competing interests between the developing team, which favors a quick rollout of the new AI assistant, and their own boss, who favors a cautionary approach. They are also introduced to the game's feedback system, which immediately evaluates players' decisions in terms of whether they demonstrate technical expertise, leadership qualities, and responsible attitudes. The purpose of this feedback system is to reward players for paying good attention to the technical and ethical information presented in the game, and to give them a reason to evaluate their different options carefully. Figure 1 shows how the feedback system (top left) is always present for players.

2. In the second level, players get to know the AI assistant and learn about its development. The goal of this level is for learners to understand, how machine-learning applications are developed and tested, and to give them an impression, how AI can help to solve complex business problems. Players can ask numerous questions and identify several problems if they are persistent. Overall, they will usually finish the second level with a positive impression of the AI assistant.

3. In the third level, players undertake the actual investigation. They appoint their two assistants to a wide variety of quality assessments, for which they barely have enough time. This time constraint makes it necessary for players to prioritize quality aspects and to consider, what kinds of risks could be involved. Through this investigation, players will usually learn a lot about the ways in which machine-learning applications can be dysfunctional and how these kinds of risks can be assessed. If they act wisely, players will identify several serious issues.

4. In the final level, players need to decide whether the AI assistant can be rolled out immediately, or whether improvements are needed. If they ask for improvements, they will need to defend them against the resistance of the developing team. The main goal of this level is to provide players experience in speaking up and addressing ethical issues at work. The resistance, which they meet, is designed to be realistic. Players who have decided to address serious issues in the AI assistant's design frequently decide to abandon the improvements they have asked for. Depending on whether players effectively demand remedies to two major issues in the AI assistant's performance, the game will end in tragedy, with mediocre results, or with fantastic business results. The scandals that can occur due to ethical problems with the AI assistant have two main functions. First, they raise learners' awareness to the ways, in which AI applications can harm business stakeholders. Secondly, they alert players to the risk that they could allow problematic products and services to enter the market if they do not secure their quality diligently.

4 Teaching Experience with CO-BOLD

Up until now, we have tested CO-BOLD in 6 courses with more than a total of 250 students of Business Administration and Information Science at Leuphana. In general, students have indicated that they enjoyed the experience of playing CO-BOLD for roughly 2.5 hours. In the initial phases, their feedback helped to improve the game significantly. The game is now perceived as mentally challenging, but students suggest that it conveys the complex concepts in an accessible manner. The game is combined with exercises and lectures as part of a short course on AI Ethics, and the remainder of the course has also found students' approval. Anonymous feedback from Management students found the game to be relatively realistic, rewarding, immersive, and educative (see Figure 2).

In terms of learning, the short course also appears to be effective. As part of their recent exams, 95 of the previously mentioned Management students (re-)took our sensitivity test on ethical risks of artificial intelligence. On average, they now identified four of seven ethical problems, whereby their total scores more than doubled (yielding a large effect in statistical terms). Obviously, we must assume that the exams provided additional motivation to study the relevant problems in detail. Further experiments will show, how much people learn from playing the game alone, or as part of more extensive courses. For now, we are highly satisfied with the outcomes of our project, since the course seems to stimulate the kind of learning and thinking that we were hoping for.

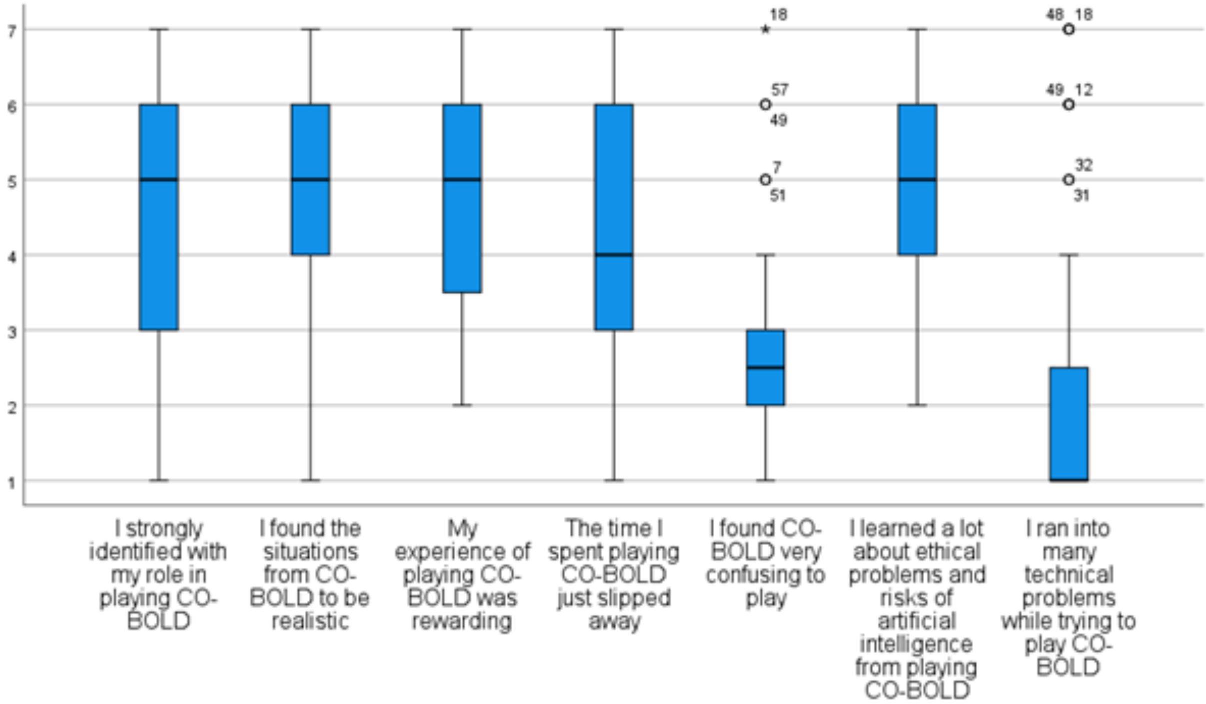

Figure 2: Evaluations of CO-BOLD from 48 Management students on a scale from 1 (totally disagree) to 7 (totally agree),

Reference

- *(1) O’Neil, C. (2016). Weapons of Math Destruction. How Big Data Increases Inequality and Threatens Democracy. Crown Books.*
- *(2) Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, E., & Vayena, E. (2018). AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations. Minds and Machines, 28(4), 689-707.*
- *(3) Coeckelbergh, M. (2020). AI Ethics. MIT Press.*
- *(4) Medeiros, K. E., Watts, L. L., Mulhearn, T. J., Steele, L. M., Mumford, M. D., & Connelly, S (2017). What is Working, What is Not, and What We Need to Know: a Meta-Analytic Review of Business Ethics Instruction. Journal of Academic Ethics, 15, 245-275.*
- *(5) Blasi, A. (1980). Bridging Moral Cognition and Moral Action: A Critical Review of the Literature. Psychological Bulletin, 88(1), 1–45.*
- *(6) Tanner, C., & Christen, M. (2014). Moral Intelligence – A Framework for Understanding Moral Competences. In M. Christen, J. Fischer, M. Huppenbauer, C. Tanner, & C. van Schaik (Eds.), Empirically Informed Ethics (pp. 119–136). Springer.*
- *(7) Tamboli, A. (2019). Keeping Your AI Under Control, Chapter 4: Evaluating Risks of the AI Solution (pp. 31-42). Apress.*
- *(8) Hurtt, K. (2010). Development of a Scale to Measure Professional Skepticism. Auditing: A Journal of Practice &Theory, 29(1), 149–171*
- *(9) Gentile, M. (2010). Giving Voice to Values. How to Speak Your Mind When You Know What’s Right. Yale University Press.*
- *(10) Katsarov, J., Andorno, R., Krom, A., & Van den Hoven, M. (2022). Effective Strategies for Research Integrity Training – A Meta-Analysis. Educational Psychology Review, 34, 935–955.*