Ethical Issues in Continuous Experimentation

PhD student: Milene Elizabeth Rigolin Ferreira Lopes[1] [2],
Advisor: Breno Bernard Nicolau de França[1]

[1]Institute of Computing – University of Campinas (UNICAMP)
Campinas – SP – Brazil

[2]Federal Institute of São Paulo (IFSP)
Campinas – SP – Brazil

m210492@dac.unicamp.br, bfranca@unicamp.br
Course: PhD; Start: 08/2021; Estimated end: 08/2025; Qualifying: 10/2023

Abstract. *Continuous experimentation (CE) is a software practice used in organizations to verify the feasibility of a product and its features. It provides procedures to verify the effect of a change by comparing different variants of a product or its features with the original version. Usually, it works as a series of online experiments with final users as participants, i.e., people with rights and duties, which experimenters need to take care of. This research aims to provide a process for addressing ethical issues in CE to support experts in designing experiments while providing ethically acceptable decisions. This paper brings reflection on ethical issues in the existing related literature and discusses them in the context of CE. Link: https://youtu.be/t53y3DdcJ3M*

1. Introduction

Online controlled experiments, such as A/B tests, are frequently used by Amazon, Microsoft, eBay, Meta, Google, Yahoo, and many other organizations to inform data-driven decisions [Kohavi et al. 2020]. Usually, it works as experiments under the application infrastructure with final users as participants, who have rights and duties. Organizations use experiments for many reasons, such as to increase product value to their users, understand user behavior, increase customer numbers, and improve their content [Yaman et al. 2017].

Although such experiments provide benefits in terms of business and product roadmaps, some organizations are putting ethical concerns aside when applying online experimentation [Benbunan-Fich 2017]. This attitude brings negative consequences to users involved in the experiment, i.e., emotional distress, financial loss, discrimination, and manipulation for financial gain [Kohavi et al. 2020]. The users are not informed about their participation or warned of possible consequences. The lack of discussion on this matter negatively impacts users' rights, and these issues remain unsolved or taken as irrelevant.

There are two exemplar cases [Benbunan-Fich 2017] [Kohavi et al. 2020] in which ethics was overlooked. Facebook's experiment, named "analyzing emotional contagion via social media" [Kramer et al. 2014], exposed randomly selected participants to more negative posts and other users to more positive ones. The aim was to verify if the participants seeing these posts would reinforce those emotional characteristics and publish negative or positive posts [Kohavi et al. 2020]. OKCupid's mismatching experiment, named "the power of suggestion", took pairs of users determined to be poor matches by the

matching algorithm (below 30% compatibility level) and adjusted the level to indicate a 90% match, indicating an excellent fit. When they believed their matches were compatible, misled users sent more first messages, confirming the hypothesis, and exchanged more messages with their supposed partners [Benbunan-Fich 2017].

These experiments do not warn users in both cases. Intentional deception was used in both Facebook's and OKCupid's emotional experiments and could be harmful to people who are already emotionally fragile. Many individuals often use social networks for friend support and conversation. Similarly, many who register on dating sites like OkCupid search for friendship and, ultimately, committed relationships. Users who are unhappy, lonely, or emotionally vulnerable, thus, may be harmed disproportionately if information in their feed or compatibility matches is manipulated [Benbunan-Fich 2017].

As these cases are not unusual, there is a need for discussing ethical issues not only in Online Experimentation as discrete instances but also in the Continuous Experimentation (CE) process, which consists of conducting experiments in iterations during the software engineering process [Ros and Runeson 2018]. This discussion about ethics should address practices such as informed consent and how professionals (experimenters) deal with ethical factors such as transparency, honesty, justice, independence, well-being, and others.

In this work, we address the lack of ethical guidance in the execution of online experimentation, highlighting the relevance of ethical issues to protect the rights of participants and showing that, scientifically, the problem stems from the need for reliable methods, approaches, or tools supporting the awareness of ethical aspects during the online experimentation process, as well as preventing unethical practices to take place. In this way, the main problem to be addressed in this work is how to deal with ethical issues in the CE to overcome the lack of ethical guidance.

To tackle this problem, we state the following research question: *what ethical considerations and practices should be implemented to ensure a continuous experimentation process that is ethically sound and responsible in software development?* This way, we aim to provide a process for addressing ethical issues in CE to support experts in designing experiments and, thus, provide ethically acceptable experiments.

Therefore, to achieve our goal, we examined ethical factors from various fields related to online experimentation. By integrating elements from different areas and aligning them with the perspective of CE, the work provides a comprehensive understanding of the ethical considerations that need to be addressed in this particular context.

Our planned contributions include (i) an initial overview of how Brazilian organizations treat ethical factors through survey results; (ii) a re-reading of the ethical factors from other areas for continuous experimentation; (iii) a list of ethical factors for CE; (iv) consolidation of techniques to predict ethical factors in CE; and (v) a process for addressing ethical issues at CE.

The work is organized as follows. Section 2 presents the background with the main concepts for a better understanding of this work, such as continuous experimentation and ethics. Section 3 presents the methods for carrying out the research. Section 4 presents the related works. Section 5 brings the preliminary results. Section 6 presents the solution proposal design, and finally Section 7 presents the final remarks of the article.

2. Background

2.1. Online Experimentation

Online experimentation involves controlled experiments, typically using digital platforms, websites, or software applications. Software product users are persistently randomly assigned to different variants, such as different product interface designs. It often involves techniques such as A/B testing, where different versions of an element are shown to different groups of users to compare their performance or effectiveness [Fabijan et al. 2020].

In software development, online experimentation supports decision-making, answering questions such as (1) Does the product or a feature solve the user's real problems and thus provide value? (2) Which of the alternative implementations do users like best? (3) Have the customers changed their behavior? (4) Does the product (still) fit the market or a segment? [Munezero et al. 2017].

2.2. Continuous Experimentation

Continuous experimentation consists of conducting online experiments in a cycle involving methods for determining the impact of a planned modification on a software product or feature by contrasting variants with the original one. The most used CE approach is designing the experiments as A/B tests. This is accomplished by segregating and exposing users to one variant (A or B) and comparing their performance data [Auer et al. 2021].

The term "A/B testing" refers to a practice for testing a hypothesis in which the variables are deliberately varied to observe the effects. This way, experimenters use one factor with two alternatives (A and B), the control (usually a baseline), and one variation. An experiment is repeated in different trials (with different participants) since the results are unpredictable and contain variance. In this design, controlled variable settings are randomly assigned to the participants [Auer et al. 2021]. In this practice, business and development processes are oriented by constantly conducting experiments and collecting user feedback. It allows empirical evaluation of their capabilities (e.g., features and quality levels) to avoid unnecessary product risks [Munezero et al. 2017].

A continuous experimentation cycle linked with decision-making was presented by [Munezero et al. 2017]. The cycle starts by identifying and selecting an idea to validate. Then, the idea is broken into assumptions. The most relevant hypothesis is chosen for a more thorough, systematically designed experiment, in which the hypothesis becomes testable. Then, the experiment is run, and, in the end, it is possible to analyze the outcomes and decide how successful the idea was.

Continuous experimentation is gaining momentum, but there are several key characteristics contributing to its success. First, it supports software systems engineering by providing a framework for iterative improvement and adaptation. Defining hypotheses and research goals ensures clear objectives are established. Metrics are used to assess user acceptance or rejection of hypotheses, enabling data-driven decision-making. Second, it requires continuous delivery of new features and updates, allowing for rapid feedback and validation. Third, it emphasizes the build-measure-learn cycle, based on the lean startup methodology, to drive innovation and value creation. Fourth, it fosters collaboration between potential users and development teams to validate proposed solutions and discover real needs. By prioritizing user data, CE enables the implementation of specific requirements or features with a high potential for user satisfaction. Experimentation strategies

such as A/B tests, Beta Tests, and Canary Releases are central in prioritizing, eliciting, and validating software requirements. Lastly, they guide software development processes, ensuring continuous improvement and adaptation based on user data and feedback. In this context, learning about customers, users, and the market takes precedence over the code itself, emphasizing the importance of customer-centricity [Erthal et al. 2023].

The term "C/D experimentation" introduced by [Benbunan-Fich 2017] refers to deceptive experiments involving programmed changes to manipulate results about users' information to intentionally mislead or misguide final users without notifying them when they are participating in the experiment. There are three aspects of ethical issues concerning C/D experiments: (i) the existence of deception; (ii) lack of protection for human participants; and (iii) lack of user agreement to participate in the experiment [Benbunan-Fich 2017].

2.3. Ethics

According to dictionaries, ethics consists of (1) "*the study of what is morally right and what is not*"[Cambridge 2023]. (2) "*moral rules or principles of behavior for deciding what is right and wrong*"[Longman 2012]. Expanding on this understanding, The definition of ethics by [Sidgwick 2011] (3) "*is a systematic and precise study of determining what individuals should do or what is morally right for them to do. It is a rational procedure that seeks to determine principles of conduct and the ultimate end of reasonable human action"* [Singer 2011].

The definition of ethics by [Singer 2011] (4) is the study of moral judgments, ethical problems, and upholding ethical standards. It involves looking into the nature of ethical decisions, identifying moral and practical judgments, and comprehending the norms or principles that serve as one's compass. Evaluation of people's compliance with ethical standards, which may deviate from social norms, is another aspect of ethics. They emphasize the use of reason in determining what is ethically right and wrong while acknowledging various viewpoints and ethical convictions.

Finally, the concept of ethics brought by [Gray and Webb 2020] (4) "*is a branch of philosophy that addresses questions about morality, such as what is the fundamental nature of morality, and how moral values are determined.*".

2.3.1. Ethics in Software Engineering

There are three ways to approach software engineering ethics. First, it can be used to characterize the actions of software engineers who make practical decisions that greatly impact others. Second, it can be used to define a set of rules, directives, or ethical imperatives that direct or mandate legislative action. Third, it can refer to a field of study that examines the connections between the other two meanings of ethics [Gotterbarn 2017].

Since software influences the lives of billions of people, some organizations have made their codes of ethics, such as the Association for Computing Machinery (ACM), the Australian Computer Society, the British Computer Society (BCS), the IEEE Computer Society (IEEE-CS), and the New Zealand Computer Society have all refined their Codes of Ethics [Gotterbarn 2017].

2.3.2. Common principles and challenges

While there is overlap in ethical principles in these areas, there may also be unique considerations specific to each area. The convergence of ethical principles in these topics is evident in the emphasis on transparency, responsibility, justice, and respect for the rights of individuals.

Challenges regarding ethical factors in online experimentation are often not adequately addressed in discrete and occasional experiments. Furthermore, as online experimentation evolves into a continuous model, the impact of these ethical issues becomes even more significant. The continuous nature of online experimentation implies that participants are constantly exposed to interventions and changes in the digital environment [Erthal et al. 2023]. Therefore, we need to understand the concepts and definitions that involve experimentation to advance the discussion toward continuous experimentation.

3. Research Method

We adopt a mixed-method approach to achieve our main research goal, encompassing:

- A literature review on CE to identify key concepts and related work;
- A literature review for investigating the treatment of ethics in related areas: Information Technology, AI/ML (machine learning), Software Engineering, and Continuous Experimentation.
- Conduct a survey with Brazilian organizations to understand how they address ethical issues when experimenting online.
- Compiling the ethical problems from the CE area and the survey.
- Propose a process to handle the ethical issues, with techniques and/or solutions already known in the literature or propose new solutions.
- Case studies to evaluate the proposed process.

The current state of our work is on proposing strategies to handle ethical issues, with existing techniques from the literature or proposing new ones.

4. Related Work

We conducted a mapping study on ethics in continuous experimentation. However, the studies found barely mention ethical issues. Therefore, we conducted an *ad-hoc* literature review on ethics in related areas (some of them presented in Section 2.3), searching for works discussing ethical factors in their area and possible ways of dealing with them. The main intersection of these works with our proposal is that they all aim to identify and find ways to deal with ethical factors within their areas. The three main related works are described below.

Yu [Yu 2020] examines ethical issues in a multi-case study and assesses their relevance to software engineering case studies. The study reviewed the literature to map ethical values, components, and characteristics. Then, 21 suggested solutions to reduce ethical concerns were introduced based on experience. Most of the suggested actions may be adapted to online experimentation, as they discuss maintaining confidentiality, ensuring security precautions are in place, maintaining anonymity, and clearly outlining any associated risks.

Cerqueira [de Cerqueira 2021] proposed an online guide, called RE4AI Ethical Guide, to assist Product Owners and developers of Artificial Intelligence (AI)-based systems in eliciting ethical requirements. To comprehend the issue, thorough literature research was conducted using the Design Science Research methodology. The manual emphasizes 11 crucial ethical concepts and is based on the ECCOLA technique, serving as a game of planning. A poll of 40 undergraduate and graduate students and a focus group with five seasoned professionals in the field were used to validate the guide.

Yaman *et al.* [Yaman et al. 2017] explore the importance of understanding ethical considerations associated with experimentation as a company development strategy. Specifically, they focus on the need to notify users when they engage in an experiment. They surveyed four software companies, including employees from different profiles, to share their perceptions and attitudes about ethical scenarios.

5. Results

5.1. Ethical Factors in the CE Perspective

As the literature on online and CE does not establish ethical principles, factors, and practices, we discuss the ethical factors established in other areas (Transparency, Honesty, Privacy and Confidentiality, Freedom, Social and environmental well-being, Justice and Equity, Sustainability, Dignity and Solidarity) from the perspective of online and continuous experimentation, according to the characteristics presented in Section 2.2. As an example, we present one factor in the following under the CE perspective.

Transparency is essential in online experimentation as it promotes accountability, openness, and trustworthiness, ensuring that experiments are conducted ethically and responsibly. Several approaches can address these factors effectively, including clear communication, disclosure of non-sensitive information, open access, and exposure to the ethical review criteria. The CE features related to transparency are supporting software systems engineering, continuous delivery to the end-users, monitoring metrics to assess the acceptance or rejection of hypotheses based on actual use, and the basis of the lean startup that involves the build-measure-learn cycle.

5.2. Survey

We conducted a survey to verify if and how Brazilian organizations consider the ethical issues inherent to online experiments and to identify which ethical issues have been considered. The target audience is online experimentation practitioners in Brazilian software development organizations applying online experiments.

The questions expected to be answered were whether Brazilian organizations understand as relevant to consider ethics when running online experiments; what ethical factors are considered and are most relevant when carrying out continuous experimentation in Brazilian organizations; and whether Brazilian organizations believe that considering ethics can increase product delivery time.

The questionnaire is composed of the following sections: (i) characterization of the participant and his/her organization; (ii) relevance list of ethical issues; (iii) ethical issues relevant to the organization; and, finally, (iv) impressions about two well-known online experiments that ethics were overlooked.

Experts assessed ethical factors and other issues using a six-point Likert scale, with one being 'never' or 'not at all' and six being 'very high' or 'always.' Also, we used descriptive statistics for quantitative analysis, and charts and numerical summaries were generated to analyze the survey data. For the qualitative analysis, coding was used, individually analyzing each response.

A total of eleven responses were received. Experts believe that addressing ethical issues in an online experiment is relevant, with a median of 6 and a range of 2. Moreover, analyzing factors per organization size, transparency, honesty, privacy, confidentiality, social justice, dignity, and solidarity did not vary significantly with company size. The average company's sustainability varies with 'little', 'high,' and 'very high' importance.

Experts believe the organizations they work with take ethical issues into account; the median is 5, and the range is 2. Most consider confidentiality, privacy, and transparency; the not observed factors are social and environmental well-being, sustainability, and dignity.

Organizations understand that handling ethics increases delivery time. Only one organization believes it does not increase the deadline.

6. An Ethics-Aware Continuous Experimentation Process

The proposed solution will be based on the identified ethical problems, proposing ways of dealing with these problems within the CE process, taking into account ready-made solutions or suggestions for new solutions, and carrying out a case study in organizations to evaluate the proposed process.

In this way, the process will take as input the design of the experiment with the proposed intervention and variables involved and have four steps: (1) design analysis, identify which ethical issues are involved; (2) selection of one or more solutions; (3) implementation of solutions; (4) solution evaluation, evaluate the solution after running the experiment, if the solution is ok, proceed; if not, realign the solution.

The main users of such a process are the experimenters, which may include managers, analysts, or developers who conduct experiments in software organizations. They would use the process during the experiment planning since it requires identifying the target audience and experiment design, including intervention and variables involved.

The intended benefits include (i) clarity on the points of the experiment that need attention related to ethical issues, (ii) the guarantee of ethical rights for the end-user, and (iii) the organization fulfilling its duty regarding ethics in experiments with real users.

7. Final Remarks

The survey results reinforce the importance of discussing ethics in online software experimentation. One can also perceive the urgency in which we need to consider some factors in this area and the need for clarification and precise ways to identify and conduct these ethical issues within organizations.

The ethical factors that Brazilian organizations consider most important are confidentiality, privacy, and transparency. A re-reading of the ethical factors of other areas in the online experimentation was carried out, as it can be noticed, analyzing the data and

the factors that are not taken into account or have a low evaluation, that one of the causes of not being taken into account would not understand the role of that factor in an online experiment.

Referências

Auer, F., Ros, R., Kaltenbrunner, L., Runeson, P., and Felderer, M. (2021). Controlled experimentation in continuous experimentation: Knowledge and challenges. *Information and Software Technology*.

Benbunan-Fich, R. (2017). The ethics of online research with unsuspecting users: From a/b testing to c/d experimentation. *Research Ethics 13*.

Cambridge (2023). *CAMBRIDGE Dictionaries Online*. Cambridge University Press.

de Cerqueira, J. A. S. (2021). Exploring ethical requirements elicitation for applications in the context of ai. Master's thesis, Universidade de Brasilia.

Erthal, V. M., de Souza, B. P., dos Santos, P. S. M., and Travassos, G. H. (2023). Characterization of continuous experimentation in software engineering: Expressions, models, and strategies. *Science of Computer Programming*.

Fabijan, A., Dmitriev, P., Holmstrom Olsson, H., and Bosch, J. (2020). The online controlled experiment lifecycle. *IEEE Software*.

Gotterbarn, D. (2017). Software engineering ethics. *Enc. of Software Engineering*.

Gray, M. and Webb, S. A. (2020). *Ethics and Value Perspectives in Social Work.* Palgrave Macmillan.

Kohavi, R., Tang, D., and Xu, Y. (2020). *Trustworthy Online Controlled Experiments: A Practical Guide to A/B Testing*. Cambridge University Press.

Kramer, A. D., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. *Proceedings of the National Academy of Sciences*.

Longman (2012). *Longman Dictionary of Contemporary English*. Pearson.

Munezero, M., Yaman, S., Fagerholm, F., Kettunen, P., Mäenpää, H., Mäkinen, S., Tiihonen, J., Riungu-Kalliosaari, L., Tuovinen, A.-P., Oivo, M., Münch, J., and Männistö, T. (2017). *Continuous Experimentation Cookbook An introduction to systematic experimentation for software-intensive businesses*. DIMECC Oy Korkeakoulunkatu.

Ros, R. and Runeson, P. (2018). Continuous experimentation and a/b testing: A mapping study. *ACM/IEEE 4th International Workshop on Rapid Continuous Software Engineering*.

Sidgwick, H. (2011). *Methods of Ethics*. Good Press.

Singer, P. (2011). *Practical Ethics*. Cambridge University Press.

Yaman, S., Fagerholm, F., Munezero, M., Mäenpää, H., and Männistö, T. (2017). Notifying and involving users in experimentation: Ethical perceptions of software practitioners.

Yu, L. (2020). Ethical considerations in case studies. *Workshop on Ethics in Requirements Engineering Research and Practice (REthics), IEEE*.