Ethics of Canary Testing: Users as Guinea pigs

Fredrik Svanholm - svanholm@kth.se

May 15, 2022

1 Introduction

Canary testing or Canary release is a method or software deployment where a new version or feature is rolled out gradually to users and evaluated over time to mitigate negative effects and/or bugs. Canary testing targets specific user groups and can thus be very powerful in optimizing software for targeted demographics and user groups. When developers are presented with the ability to target things towards specific users certain ethical concerns presents themselves. When the power to target users and to evaluate how they respond to different things appear developers may run the risk of engaging in ethically questionable practices. User experience may be negatively impacted due to rapidly changing layouts and functionality of software in order to analyze and maximize for company profits, retention and engagement. By keeping testing practices transparent with users to keep the team accountable for their testing practices along with trying to make sure the interests of the users are kept in mind and prioritized and giving users the autonomy of opting in our out of these Canary tests, developers can mitigate the risks of ethically questionable practices while still utilizing the flexibility and power of canary testing.

2 What is Canary Testing

Canary testing is a method used to release, test and evaluate new versions of a software or a service in the field of Software Engineering. The name Canary testing comes from that Coal miners used to bring Canary birds with them down into the mines in cages and if the canary birds died they knew it was a sign that lethal gasses were too high and that it was time to evacuate and leave the mines. [2] Similarly in software testing, Canary tests are used to prevent large scale catastrophes and complete system failures when rolling out new updates.

The idea behind Canary testing is similar to Rolling deployment where specific servers in the live production environment are given a new update and the results of the new update are evaluated and gradually rolled out to more nodes in the server network if results seem to indicate an improvement or that it works. But with Canary testing instead of deploying the update to specific servers, the new updates are instead rolled out to specific users of the application. the new version of the service to a small subset of users of the application or service and if evaluation metrics indicate that the update is working or is an improvement if it points towards an improvement then the new update continues to roll and to a larger subset of users with the goal of it eventually reaching all users [1]. For a visual representation of the gradual deployment see Figure 1.

3 Benefits of Canary Testing

Canary testing can allow developers to test new versions without the risks of exposing an entire user base to uncaught errors or bugs in new updates. If developers were to roll out their new updates to all of their servers or users, then that risks having the entire system or all users being exposed to potentially harmful uncaught issues or errors in due to the update. By having some sort of Gradual deployment of new updates developers can mitigate the risks of colossal total system failure and minimize damage of uncaught errors. There are different ways of gradually rolling out new updates

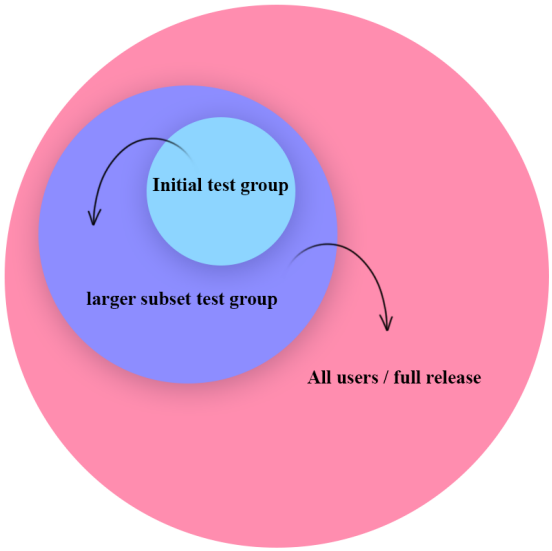

Figure 1: Gradual deployment to users with Canary testing

other than Canary testing such as Rolling Deployment and Blue/Green deployment [1].

Canary testing allows developers to gradually roll out and evaluate updates and or new features of a service to a small subset of the user group. Compared to Rolling Deployment developers can target these new test builds to specific sub-groups of users. By doing this developers can test out the update in a live production environment and get real user test data while not risking the entire service going down. Targeting specific user groups can also have certain benefits over Rolling deployment since developers have control over which users that will get a certain update. Developers can deploy new test builds with certain features that they think may be useful for a certain subset of users that are more likely to interact with a given new feature and thus get more test data per user.

This concept of target group evaluation can be extended even further to optimize and compare how changes of an update effects different target demographics, and through this gained knowledge optimize the app to adapt and better suit each users needs and preferences.

4 Potential Ethical Concerns

One of the core ethical risks when running Canary tests are what is also considered one of the big benefits of Canary testing. What differentiates Canary testing from other gradual deployment methods is it's focus on rolling out updates to specific users. If developers are not careful in how they utilize this power they may run the risk of engaging in ethically dubious practices.

As Helena Jeret-Mäe writes in her article *The 10 Commandments for ethical software testers* [**?**], it's very important to know which interests that your testing serves. With this power of targeting user groups developers may run the risk of potentially running tests on users that does not serve the interest of the user. With the ability to target and test releases on specific groups, investors and stakeholders in a service may push developers to experiment with features that target certain demographics to optimize profits, retention and or other metrics. While Canary testing can be used to improve the user experience and to quickly roll out updates while mitigating potential damage, it also has the potential to be used against the interests of the users, which are the ones that will be effected by the tests.

Developers may be rolling out a new test version to users without properly evaluating the changes in the service, how it may affect the users, their ability to use the service and the reason behind the test. It's the users that will be effected when performing Canary tests and they are the ones running the risk of running into issues when a new update rolls out. Users may be able to stand behind and accept

running into potential bugs and changes with a service if they feel like it is done for the betterment of their experience and for their own long term benefit. But if users view a test or a new feature test as mostly or only being done in the self interest of the company it can at best create resentment and negative views towards the company and at worst may be seen as ethically dubious or wrong by the user base.

Rolling out updates to a subset of users, if not done carefully. Can give viewers a sense of loss of autonomy, especially in the case of a payed service or licensed product where a certain level of quality is expected. When customers pay for a service they expect it to work and for it to be stable. If users are not aware of that they are being tested on they may get frustrated when they realize that their version looks or behaves differently from their peers versions or from what they find written about the service on the internet. Rolling out tests to users without their knowledge has the potential to harm their ability to utilize a tool that they've payed for. A lack of transparency in how these tests are performed and how the user base is being used may lead to users miss trust and negative brand view.

5 Possible Mitigations

When engaging in Canary Testing it is important to always evaluate how the tests are being performed. One way of making sure that ones Canary tests are ethically sound is to evaluate who's interests that the tests actually serve. Does it benefit stakeholders and investors by manipulating user behavior or does the test provide a new features set that may aid in improving user work efficiency and well being? This can be done by correlating each of the new things that are being tested to company profits, users or maybe both. Most companies are for profit organisations and of course companies need to make monkey. But if a tests seems to be too heavily weighted towards serving the companies interest without giving much of any benefit to the user base, then maybe rolling out Canary tests isn't the right solution.

Developers may not always be in a position where they have control over how Canary tests are being used. They may be forced by upper management to do or perform certain test practices that may be viewed as ethically dubious. In such cases developer can create frameworks to evaluate tests prior to deployment. These frameworks can have user benefits be considered and weighed against the possible user experience impact. By making sure to minimize user impact when engaging in test that does not serve their interest it is possible to mitigate possible negative reactions from users. If a new version ready for canary testing is considered to be in a risk zone or potentially not beneficial for users or that it risks impacting their experience negatively while not serving their interest, then adjusting the test version may be in order or it may be good to suggest another way of evaluating the new features in order to avoid engaging in ethically questionable practices.

Another way of mitigating risks of negatively impacting user relations is to give the user base some autonomy in their participation of the Canary tests. Developers can incentivise users to participate in testing by allowing them to opt into special pre-release builds specifically for testing. An example of users having autonomy in their participation of Canary tests can be found in how 343 Industries, the developers in charge of the game franchise Halo, has has an opt in play-testing option for players to test out new things with the knowledge of it possibly having bugs and issues called "The Halo Insider Program"[3]. Having an inner circle of testers which is opt in allows users to have the option to opt in or opt out of testing new features. It allows users to evaluate if it's worth the risk of bugs to gain access to new features faster. Some power user may find it worth it while some more casual users may not have a desire to get the newest things as fast as possible and values a stable experience higher. By lending users some autonomy they will feel more in control over their experience and if a user participating a Canary test runs into issues they will have the autonomy of opting out.

It can also be a good practice to be transparent about the tests and the metrics used in the Canary tests, that way developers can be scrutinized and held accountable for their testing practices by their users and if some practices are ethically questionable the pressure from a user base may be helpful if

developers wishes to change things in their testing practices when discussing it with upper management.

Alternatively one could introduce the ability for users to opt out of canary tests rather than to opt in. A key aspect in making ethically sound Canary tests is to give users some agency. If users are negatively effected by a test that they have no control over, they will likely project their frustration towards the company, provider and or developers of the service. It's important to make the users they feel like they are in control of their user experience and if they feel like they are negatively impacted by canary tests that they have the ability to opt out. Prioritizing user agency means that users can do something about their frustration if they do not approve of a companies testing, practices or how they are effected by it and thus instead of projecting their negative feelings towards the developers they can do something about it.

6 Conclusion

Canary testing allows developers to quickly iterate and test new features without effecting the entire user base. But With the power of targeted testing on sub groups of users developers run the risk of engaging in ethically questionable practices if Canary testing is not approached and handled with caution. There are three main things developers can do to try and run ethically sound Canary tests:

- Evaluate who's interests that the test serves and how the tests can effect the users.
- Be transperant about how the tests are carried out and what methods and metrics that are used to evalute the tests.
- Give users the autonomy to opt in or out of testing.

Developers must make efforts to evaluate who's interests that a tests serves and how it may effects it's users. Developers should make efforts to disclose how they run their tests and what metrics are evaluated to keep themselves accountable by the users in case they are blinded from their own actions. And developers should prioritize user agency in their choice of participation in Canary tests. Doing all of these things will aid developers in creating ethically sound Canary tests and allow developers to avoid engaging in ethically dubious practices when utelizing the power of Canary testing.

References

[1] Ali, A. Identify and minimize production risk of log4shell. *Dynatrace*.

[2] Wikipedia. Lightweight directory access protocol. *https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol*.

[3] Wikipedia. Lightweight directory access protocol. *https://en.wikipedia.org/wiki/Lightweight_Directory_Access_Protocol*.