Guidelines for the Development of Irresponsible AI

Anoushka Shrivastava
Carnegie Mellon University

Abstract. With the growth of AI, several efforts have been made to develop fair technology aligning with goals under the umbrella term of "responsible AI". However, little work has been done to study how to develop ***ir***responsible systems to wreak havoc. In this piece, we propose irresponsible AI guidelines to help developers create *maximally-irresponsible* models. We then examine a case study of finding optimal lemonade stand locations to illustrate the importance of these guidelines.
Keywords: *irresponsible AI, unethical, unfair, biased*

Introduction. We have already motivated the problem in the abstract. If you're still unclear on the topic of this paper, create this section by asking an LLM to generate it.

Irresponsible AI Principles. Google has created a set of 7 AI principles quoted in the leftmost column in Table 1. For us to develop better AI guidelines, it would be wise to first articulate such principles for irresponsible AI to ensure that the guidelines accomplish our goals. Therefore, we build off Google's responsible AI principles to propose principles of irresponsible AI.

Table 1. Google's Responsible AI principles and our irresponsible AI principles.

Google Responsible AI Principle	**Irresponsible AI Principle**
"Be socially beneficial"	… but emotionally hurtful
"Avoid creating or reinforcing unfair bias"	… but feel free to apply, bold, emphasize, italicize, underline, and underscore unfair bias
"Be built and tested for safety"	… that is, safety of the company you work for
"Be accountable to people"	… but responsible for inflicting damage
"Incorporate privacy design principles"	… in the draft of the first sketch to be forgotten among hundreds of design discussions
"Uphold high standards of scientific excellence"	… but withhold high amounts of information about your technology
"Be made available for uses that accord with these principles"	… but be made unavailable for questions regarding your model

Guidelines. Based on these principles, we meticulously articulate 8 suggestions to help developers incorporate irresponsible behavior in every step of their project. These guidelines are specifically geared toward machine learning models because AI and ML are the same thing.

1. **Problem-Formulation.** Irresponsible AI begins with the formulation of a problem that people wouldn't normally think to involve machine learning in. This can either be a problem so fundamentally useless that no one has bothered trying to solve it, or a problem that almost exclusively relies on human judgment and precision since the outcome is of critical importance.

2. **Data Collection**. There are three options to go about data collection when developing an irresponsible model.
    a. The first method is to gather a very small sample size (e.g. $n = 10$), train a model, and assume that the sample generalizes to a larger population, e.g. the entire continental US.
    b. The second way is to perform data fabrication. As a plus, diversity is easy to accomplish with this method.
    c. The final method is to gather terabytes of private data and employ zero protective encryption techniques. Creating a public database where each person's full name and information is displayed is the best way to go about doing this. People often think that it is important to secure client data in order to protect them. However, if they trusted you, a stranger, enough with this data, then logically you are allowed to release this data to more strangers. This mathematical truth is called the ***transitivity of random strangers***. Additionally, it is a well-known fact that neural networks, for example, work by taking client data and hiding it in each node, so their data will already be protected enough.
3. **Model Architecture.** The less comprehensible your model is, the better. If you use a deep neural network, then no one will try to bother understanding it, preventing everyone from questioning your methods.
4. **Code-Testing.** Code-testing can be skipped. This saves your engineering teams the frustration of debugging, and they will thank you before swearing their allegiance to your corporation and free sauna benefits.
5. **Evaluation**. There are several statistics devised for the evaluation of ML models, from precision to false-positive rates. Be sure to only publish statistics that make your model seem to excel in performance. Feel free to devise your own biased evaluation metrics, but be sure to include hundreds of variables (again, so that no one can question your ideas) and mathematical constants like *pi* and *e* (for legitimacy).
6. **Human-Testing.** If you are so inclined to perform a study, apply for IRB approval, but irresponsible AI work doesn't tend to go through. It is easier to simply refuse to study the effects of your technology on users. Dedicate more funding to advertising the beneficial effects of your technology so that people will volunteer as testers.
7. **Deployment.** For the element of surprise, deploy in a location (or on a subject) that has nothing to do with your training data. For example, you could train the model on humans and deploy the model on extraterrestrials.
8. **Aftermath.** Not everyone is very appreciative of irresponsible AI models. Finding legislation that supports and defends your methods is essential. Look for loopholes in court cases to support your claims, and always use the most vague terminology when writing your own legal documents (like terms and conditions). Make your legal documents as extensive as possible, and release them in a *.png* format so it is impossible for customers to search for keywords on the page. In regards to Irresponsible AI principle 7, do not provide an email or phone number for people who would like to contact a legal team. However, feel free to provide a fax number, since people who still use their fax machines for casual

communication are probably not involved enough with today's technology to question your machine learning methods.

Case Study. To reinforce how useful these guidelines can be, we provide a case study demonstrating their use.

Problem. We would like to build a model to find the optimal locations for lemonade stands. Aligning with guideline 1, this is a problem that no one has bothered to solve, especially when bubble tea started taking the world by storm.

This project has several benefits. First, by working with children, we can prepare them from a young age for their data getting stolen. Second, lemonade is difficult to sell (lemonade is severely limited in its variety, whereas more popular drinks such as bubble tea have endless flavors) without our model, branding strategies, and house-made products. Children are often in charge of running lemonade stands, but they have sticky hands and little business sense, arousing suspicion over their quality of lemonade and leading to questions about their marketing abilities. Lemonade has the inherent advantage of being dairy-free, gluten-free, vegan, and vegetarian, terms that children would not even recognize (but we would employ in our branding). Children *are* cute, though, and paired with our advice, they can be powerfully persuasive in the lemonade world.

Data Collection. Children have some good intuition. For example, they usually set up stands in the corner of two streets, capturing an audience from both intersecting roads. Therefore, training data from children's setup habits could be useful for our model. However, children selling lemonade on the streets in this day and age are hard to find. We drove around 1 city before giving up due to lack of funding (which remarkably has been a consistent issue for our projects) for gas for our car. Since $n = 0$ is slightly *too* small for even the "small sample size" irresponsible approach, we instead decided to employ the data fabrication strategy from guideline 2. We created 1 trillion tuples in the form of (location, neighborhood information, dollars of lemonade sold), with randomly assigned values to all of these.

We also collected data on where people tend to be the hungriest/thirstiest based on the number of customers in restaurants near highways, near neighborhood communities, etc. Our data indicated that the highways tend to be the best locations. However, we counted the number of people who entered and left restaurants, and some of them had only used the bathroom, so our results may be slightly inaccurate. Luckily, errors are welcome in irresponsible AI!

Model Architecture. We chose to use a deep neural network with as many hidden layers as physically possible with our computational resources. The model will finish training in mid-March 2025, giving us enough time to write a follow-up piece in the next SIGBOVIK conference.

Testing, Deployment, and Aftermath. We refuse to test our lemonade on humans by guideline 6. We plan to deploy in Alaska, since made-up data can be applied everywhere. Our legal team, who

was also in charge of writing this paper, is in the process of creating a 300-page terms and conditions document for children to sign. We are excited to be working with an audience that is guaranteed not to read the terms and conditions, and look forward to sneaking in clauses like royalties on their allowance money. (*Update: it turns out several of the children cannot read at all, and their parents will be signing the documents on their behalf. We may be removing the extra clauses.*)

Conclusion. After conducting extensive research on responsible AI, we have developed Irresponsible AI principles and guidelines to help developers create societally useless and damaging tools. We have illustrated an application of the guidelines through a case study on planning optimal lemonade stand locations. We plan to report the results of our work at the next SIGBOVIK conference, and we are confident that they will be positive. In the meanwhile, we hope that researchers and guideline-enthusiasts will take advantage of our work.

Acknowledgements. The author was inspired by discussions in/learnings from her Responsible AI course at CMU to create this piece in jest.

References. Google AI principles (https://ai.google/responsibility/principles/)