The Responsible Innovation Framework: A Framework for Integrating Trust and Delight into Technology Innovation

Alka Roy
The Responsible Innovation Project
alka@responsibleproject.com

Abstract

Although systematic biases in our intelligent systems and lack of privacy, equity, and ethical and trust considerations have entered AI and emerging technology debate, we are still lacking a common practice-based framework for innovation that puts social well-being if not ahead at least on par with growth and profits. This comes at a cost that includes public trust. This paper introduces The Responsible Innovation Framework as a tool with a reframing of stakeholders, value-sets, and influences. Who is this for? It's for everyone who's involved in decision-making for products and technology especially leaders and practitioners. The paper 1) makes a case for using a common framework starting from the ideation and vision stage or introducing it anywhere in the process, 2) describes the "essential" components of the framework: stakeholders, value sets, and influencers, 3) provides examples of how value sets could be leveraged in a flexible and iterative way for AI or Non-AI technology, and 4) lays out the need for additional work and case studies. The goal of the framework is to include social considerations as an essential part of technology decision making.

1. Introduction

Digital innovation has given us unimaginable access, connection, information, personalization, and convenience. But it has come with a cost. In the next 5 years, leaders in technology industries--Cloud, 5G, and AI-- will converge with automobile manufacturers and device companies, and various other startups and interface developers to transform transportation, healthcare, and education systems. These companies each have their own unique set of values and cultures. Though they are designing our future, they lack a common framework to align to build technology solutions for our public systems that allow for public concerns or input to show up in any meaningful and consistent way.

Today, only 53 percent [1] of the world are internet users. Exposed public betrayals of trust like the Cambridge Analytica scandal [2] and YouTube's child privacy violations [3] have not resulted in long-term changes. News is littered with technology and media companies harnessing digital activity to create algorithms that control and influence what people see and hear. Books after books have been published by tech insiders sounding alarms on data ethics and the imbalanced impact of technology in our society. Fake news, deepfakes, filter bubbles, and echo chambers have entered our lexicon. In 2016, Americans were collectively shocked by the news of social media being used for election meddling. Four years later, in 2020, the possibility of election meddling is accepted as reality [4].

With the fast pace and vast variety of innovation, even the insiders in tech companies are often unsure of how to evaluate new technology or its impact. The majority of the technologists I interviewed for this framework wanted to change the outcome without changing their methodology or amount of outside interference, the hope seemed to be 1) not to get things too wrong, 2) avoid getting caught if you do, and mostly 3) stop other bad actors from doing harm. The possibility that without checks and balances, well-meaning people get things wrong, was rarely mentioned. Though several companies have public or private AI principles and best practices, the priority continues to be speed, cost, and revenue potential. To assure compliance, resources may be directed, after the fact, to compliance review or audits, supply chain contract language, and PR. Band-aid solutions that might make the user experience cumbersome or

URI: https://hdl.handle.net/10125/70738
978-0-9981331-4-0

programs like "AI for good" are set up to redirect public attention and win trust.

Generally, considerations that impact human well-being are an afterthought and are seen as the cost of doing business rather than being integrated into the prioritization, process, and evaluation of innovation. What if we leveraged the flexibility and agency that is already built into technology development, where experienced and motivated engineers, product managers, and designers can make or influence key design decisions within cost and time constraints? What if there was a common visual and accessible framework that could be used starting from ideation that was flexible enough to apply to different scenarios and in differing degrees? Could it shift how we innovate?

2. Methodology Considerations

The Responsible Innovation Framework evolved after researching prior work in responsible innovation and assessing guidelines and frameworks for social innovation, responsibility, ethics, and trust in AI [5][6][7][8][9][10] as well as studying the impact of influences, framing of human & machine and behavioral models that have been extensively studied and debated [11][12]. The framework was compared and contrasted against key AI ethics guidelines [4.4, Table 1]. The fundamental questions were: Why are we designing technology the way we are designing it? What is missing or needs to evolve? Most often technology and product innovation occur within the industry or industry/academic or open-source/industry partnership within their normative and cognitive rules.[14][15] The language used in the framework was one that would translate to this world of technology practitioners and innovators. There was also an effort to create a simple and visual representation that showed the interdependency of value sets instead of a checklist approach.

In addition to 1/1 interviews, two small workshops were held with 50+ technology leaders (data scientists, AI program leads and ML engineers, trust, and AI ethic experts, product managers, UX designers). They were based in the US and Europe with an expectation of six that were based in China and India. During the workshop, the attendees were asked to apply the framework to a case study or a product of their choice on their own (results and case studies will be published separately) and then reassess after a walkthrough of the framework.

The framework is also informed by the failures and successes of my 15+ years of industry experience with emerging technology and 100+ product launches. The key learnings came from building speech apps (Conversational AI) for enterprise and retail clients while trying to build accessibility and inclusion into mainstream products and my work in Ethics, Responsible and Trusted AI with industry, academic, open source, and grassroots communities.

3. Related Work

3.1. From AI & Ethics to Innovation & Responsibility

AI Ethics and Trustworthy AI work has been important in raising awareness and debates. Often, the key frameworks tend to 1) either overemphasize or isolate the ethical lens to AI instead of applying to the wider technology 2) and create a list of requirements that lead to a culture of compliance or governance rather than integrating the complexity of considerations into ideation and technology development. At times, an ethical guideline may be incongruous with a company's culture and business model and become mere lip-service. In a 2020 study, Hagendorff compared 22 ethical guidelines to conclude that ethical guidelines often do not have an actual impact on decision-making in the field of AI and machine learning [13].

The language and framework of responsibility are a better fit for technology practitioners and innovators. "The effects of decisions or actions based on AI are often the result of countless interactions among many actors, including designers, developers, users, software, and hardware… With distributed agency comes distributed responsibility" [16].

The countless interactions are further compounded by countless vendors and systems that go into creating technology solutions that include AI. Focusing on AI Ethics too narrowly instead of technology as a whole misses out on the inherited and interwoven challenges.

3.2. The Legacy of Responsible Innovation

Borrowing from previous work, this paper redefines Responsible Innovation as innovation that invests in technology, people, and the environment today with the goal to create a delightful and trustworthy future for everyone. Responsible Innovation aspires to be human-centered and environmentally-friendly and invests in future relationships with key stakeholder groups to drive technology adoption based on delight and trust. It is

innovation for the people, by the people. It is the mindful and deliberate design of our future.

Responsible Research and Innovation (RRI) was formalized by Schomberg's thoughtful analysis as "a transparent, interactive process by which societal actors and innovators become mutually responsive to each other with a view to the (ethical) acceptability, sustainability and societal desirability of the innovation process and its marketable products (in order to allow a proper embedding of scientific and technological advances in our society)" [17]. The focus of RRI in advanced research and a framework focusing on how to bring public concerns to research was documented for a geoengineering project[18]. That study focused on policy and governance and defined Responsible innovation as taking care of the future through collective stewardship of science and innovation in the present"[18]. This paper builds on these thoughtful works through a new framework that can expand and further translate Responsible Innovation to the practice of technology and product innovation and development.

4. The Responsible Innovation Framework

To bring Responsible Innovation to technology and product innovation and development, this paper proposes a framework for practitioners, innovators, and decision-makers. The aim is to raise awareness of interconnections, overlaps, and conflicting interests that affect decisions about technology research, design, and implementation. The Responsible Innovation Framework includes three major stakeholder groups, three common value sets, and three key influences.

4.1. Stakeholders: The 3 Key Groups

A. People
B. Things
C. Environment

4.1.1. People: Innovation for Human Well-Being

Any innovation of processes, products, or technology, needs to remember who they serve—people. This includes the individual and collective needs of workers, leaders, consumers, and users.

Often, excitement about what a new or emerging technology can do or the economic value it can bring overshadows the obvious questions: Who is it for? Why do they need it?

This oversight is not always malicious but rather habitual, influenced by an implicit business culture driven primarily by financial metrics. In other words, tech makers are not rewarded when they put people first or a wider group of people, especially if those people are a vulnerable or unsuspecting group or not typically marketed demographic. That will need to change for products, businesses, and technology (including AI systems) to include a wider group of people and their well-being as a key stakeholder.

There are so many examples, even as leaders at tech companies talk about inclusion and users first, of predictive models, surveillance sensors, or companies selling their customer's online behavior and data to third party companies. For example, is it acceptable to record children and broadcast their images and activities on the public internet when they are too young to understand the implications and consequences? Should tracking, content discovery, and ownership of data be different for children versus adults? Should we wait for regulations or invest in offering better protection for children online?

Though the assessment of trade offs of competing interests of individual and collective well-being of different impacted groups can be tricky, they are not trickier than considering other trade offs. Mere consideration and transparency about these considerations can lead to wider trust-driven technology adoption.

4.1.2. Things: Systems, Products, & Technology as a Means for Innovation

Our tendency to anthropomorphize and idolize the objectivity of technology impairs our judgment. By design or subconsciously, it clouds our ability to leverage "things" accurately and at times more effectively. We talk about trusting machines—building empathy and humility. We give things gendered or non-threatening names. Not only that, but we give machines and systems human personalities (authoritative or friendly) and characteristics (fingers or faces). Or debate whether "they" are rational or have "consciousness" or will "save" us from ourselves. There is no "they" without "us." This confusion, though natural and age-old, has far-reaching implications.

This topic warrants deeper and separate exploration including revisiting one of the general interpretations based on Turing's Imitation Game [11] where he asked, "Can Machines Think?" which has led us to the gauge mimicking humans beyond recognition

as one of the key success criteria of an intelligent machine (AI). But is that ever a fair comparison? One person being tricked by a large set of hardware and software systems developed and operated by a team of people and organizations over time?

This flawed approach, habitually or on purpose, can make our design "things" inefficiently or use them in a way that makes us drop our guard or give up our privacy or agency at the moment and potentially regret it later. This can create confusion or fear and erode trust over time. It can also make us design things that encroach on our human agency and transfer our implicit biases and characteristics—including manipulation or lack of transparency for self-interest—at scale. The best thing about "things" is that they don't care (even when they might be designed to make decisions and say things that seem caring). Things operate according to the goals we set or design for them (even when we are unaware of the implications). They interpret our commands based on the sophistication and autonomy that *we build into them*.

The communication between systems and machines is different as it is between humans or humans and machines. It is imperative to remember that in the accelerated speed and fascination with automation or while designing autonomy. If we are to design machines that understand us, we also need to remember and be clear that they are things designed for humans. And will ultimately impact human to human interaction.

Consider the questionable gender and potential targeting of underage girls embedded in a chatbot modeled as "a virtual teenager girl." Here you have chatbots that people know to be "rational, unfeeling machines" being described as "likable" and witty" [19]. Though such chatbots may feel personal and intimate to humans interacting with them, they are not private and empathetic--they are systems. They were and are being monitored, tracked, trained, retrained, updated, and redesigned by a group of people based on behavioral and psychological research to drive usage and engagement. They are navigating networks and clouds and intelligent systems that raise questions about privacy and security.

They fail both stakeholders: people and things, even if the oversight is not meant to be malicious. In the virtual teenage AI chatbot example, the false personality does not consider the impact of exploiting a vulnerable, targeted, and underage demographic. Teenage girls are known to have issues with their body image that are exacerbated by unrealistic depictions in media already let alone having to be compared to CGI generated images [20]. The chatbot has a girl's image and is dressed in what appears to be a school uniform. What is ironic, is that this choice may even keep the product from showing its technical chops and having a wider appeal. Imagine, if the classification had non-human fictional visual characteristics and the emphasis was on its ability to understand human communication, rather than pretend to be one.

4.1.3. Environment: Time & Resources: Innovation Beyond Sustainability

Globally, we share one planet, our universe, and amazing yet finite resources of time and resources. The design criteria and best practices should guide the sustainable and regenerative use of these resources rather than a limited overuse for short-term efficiency.

This stakeholder is the most obvious but complex to navigate. This can be seen as the planet or the wider universal laws and need further developments and examples. Extensive work has been done and continues to be done to understand models, frameworks, and patterns of regeneration that emphasize the "co-evolutionary, partnered relationship between human and natural systems" [21] and needs further inquiry that is outside the scope of this paper.

Sustainability is a start, but regeneration should be the goal of innovation. Shifting to the challenge of expanding the value system while minimizing our environmental impact will bring us different results. Or even considering nature and its laws as a stakeholder allows us to expand the value chain (i.e., quantum) and open new possibilities. Nature will survive us. The question for collective innovation is whether humans can get their act together to survive themselves.

4.2. Values: 3 Common Value Sets

Based on Responsible Innovation work and research on trust and AI ethics guidelines mapping as well as reviewing best practices and design principles, the three common "essential" value sets for responsible innovation are 1) Delightful & Trustworthy 2) Dependable & Inclusive 3) Open & Safe. These values are listed in pairs because there is a need for us to consider the tradeoffs and the balance between the sets to get to an optimal solution for a particular use case. Looking at these values as interrelated can lead to a richer solution. The tradeoffs and balance between and among these value sets offer room for flexibility and interpretation based on the use case, scenarios, user base, and industries. Without this flexibility, different

technology practitioners and innovators, with different goals and business models, will not be motivated or able to leverage these values as the technology or product evolves.

4.2.1. Delightful & Trustworthy

Innovation with people in mind needs to include a balance between delight and trust. One without the other can feel lacking. Consider a magician who makes a dove appear out of nowhere or reassembles an assistant she appears to have sliced in half, but everyone knows it is a trick. The unusual creates delight. And though it is a trick, the audience is paying the magician and hopes and trusts them to do their job of tricking them well and that transparency gives the magician a license to perform their tricks without causing harm. Contrast it with a casino that benefits from tricking the players and pretending that the games and slot machines are a game of chance when they are designed to be in the favor of the casinos. They are winning at the cost of the players.

This may seem counterintuitive but in the case of the magician trust and delight are balanced because of the right amount of transparency. This question has raised much debate in the AI ethics space: what is the right amount of transparency [22]? The answer is--it depends. If someone has never seen a magic show and doesn't realize that they are watching a trick, watching an assistant cut in half would and should be horrifying (no prior knowledge or transparency). On the other hand, if the show started with disclosure and provided play by play detail of how the tricks work (complete transparency), the magic show would become not as interesting unless we were attending a class or workshop to learn magic.

Another common experience is navigating long lists of disclosures where the user needs to accept terms and conditions in order to access information for a digital service. Do the users have a real choice to get what they want, as quickly as they need or want (delight) without giving up the details (trust)? Or is the disclosure veiling transparency in inaccessible legal and lengthy language. The person's desire for immediate gratification (delight) competes with their concern for protecting their data and interests (trust). And the ultimate decision is sometimes made within a click depending on how they assess their options and how the interaction or product is designed.

4.2.2. Dependable & Inclusive

To earn trust, technology or product also needs to perform consistently and be dependable. Reliability, availability, and resiliency are the measurable basics of good engineering and design practice. And a core value of scientific rigor.

To be inclusive, the product must be available and accessible to as many impacted current and future users as possible. Compromises are made when inclusion and accessibility are seen as competing with the performance or business models. Or when it is either not realistic or enough time has not been allotted to be able to test and verify a larger set of dependencies and ensure a high level of performance and dependability.

The dilemma is that if there are no real alternatives for the users or groups that are not included, these potential users or customers go unserved, excluded. For example, one of the challenges of net neutrality was the competing interests of inclusivity (allowing everyone equal access) with dependability and performance (giving preferential quality of service and resource allocation to real-time services like video or voice vs. email or downloads).

Applying this to the previous example of Xiaoice teenager-girl-like chatbot, the considerations of the impacted group, teenagers, or those who care about the image, safety, or well-being of teenagers and women were not considered. Even though the 660M men and users find the chatbot dependable and reliable, though likely a biased, stereotypical, or inaccurate construction of a teenage girl's interaction[19].

4.2.3. Open & Safe

When it comes to the value set of openness and safety, the central question is: Given the particular application, goals, and circumstance, how open can the innovation be while being as safe as the public or its users need it to be?

This debate has been going on in the open-source community but there seems to be a disconnect between discussing ideas for a hypothetical and high-stakes technology implementation versus understanding the basic concern. Consider this real-life scenario (the author witnessed) that involved an industry open-source meeting about Trusted AI that was not public. The moderator turned on an app to record the panelists' conversation. The intent was to be open and accessible and post the transcript publicly. Why would this be an issue at an open-source forum? The team

was debating complex, controversial topics. Some members didn't feel comfortable having half-formed personal ideas recorded and posted online since they were also representing their companies. The intention of "openness" has a positive connotation, but it can inadvertently lead to a lack of safety, which then keeps certain and often vulnerable members from fully engaging and trusting. In the end, the group decided to make meeting notes about decisions public but keep the discussion portion of the transcript for attendees-only negotiating a balance between open and safe.

This negotiation can lead to graver implications violating privacy, safety, and security when data-sets without consent or clear lineage are made public and can never be fully recovered. In the case of the chatbot Xiaoice, when Microsoft licensed their AI framework [23] that had earlier created racist & offensive chatbots in the US (Tay & Zo were both shut down) [24]. These open licenses, in the wrong hands, could be used to create even worse and more exploitative adult content by users or companies.

4.3. The Three Influences

The three major influences captured in the framework that impact innovation are 1) Law & Policy, 2) Market & Economy 3) Culture & Norm. This list borrows from Larry Lessig's list of modalities [25], with one fundamental modification. In the Responsible Innovation Framework, they are considered influences instead of adversarial or constant pressures to back against. Depending on the circumstance and perspective, they can serve as limitations, pressures, or resources.

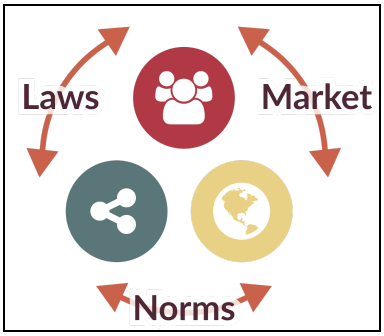

Figure 1: The Three Major influences

These influences are our current reality. This is how we have organized ourselves. They create operating boundaries for almost every decision we make as well as provide checks and balances between competing interests. If all three influences are (openly or behind the scenes) controlled by the same entity, it can result in a lack of checks and balances and a power imbalance that can lead to potential harm. And the overall system will be eschewed to primarily serve the interests of those controlling the ecosystem.

For example, an organization's culture is influenced by its company's culture and norms which, in turn, is influenced by industry realities and the countries where they operate. But if one company or industry has control of all three influences, there is an imbalance towards that industry's self-interests.

The two most common sets of beliefs shared during informal interviews were, first, that social values, rules, and regulations "are nice to have but can slow down progress." Second, that "regulation kills innovation." Both beliefs are not grounded in facts. The industry has consistently innovated around social and legal constraints. Rather, the challenge seems to be a willingness to see how the regulations can serve as an impetus for trust-based value creation for a larger group of stakeholders. Technology is grounded in rules and processes that are constantly evolving but also has a long history of standardization (even when competing companies) for common interests like mobile roaming.

It is important to explicitly call out the biggest influence on innovation today—economic drivers and market-revenue—time, cost & money. They are seen as unforgiving but also rewarding and currently provide the greatest motivation to technology companies. Self-regulation is an important aspect of innovation because regulators or community advocates are often behind the curve. But so is the market. It is often the researchers and technologists who "help" articulate or interpret the "economic value" of their research, invention, or innovation. Though redesigning economic and regulatory influences are outside the scope of this paper, they will be needed to create incentives and a lasting impact of responsible innovation.

The framework itself is designed to help with this self-regulation, the "Culture & Norms" influence, by shifting the technology-making cultural norms for innovation. Because similar to the standardization of technology, self-regulation without a common and "acceptable" framework will fail. The idea of markets, laws, and norms are complex and need further nuanced development. They are mentioned here to acknowledge their role in this framework.

4.4. The Collective and Interdependent Responsible Innovation Framework

The stakeholder groups (people, things, and nature), common value sets (delightful & trusted, dependable & inclusive, open and safe), and key influences (laws, market, and norms) are combined in a visual view of the framework for human-centered, ethical, sustainable innovation.

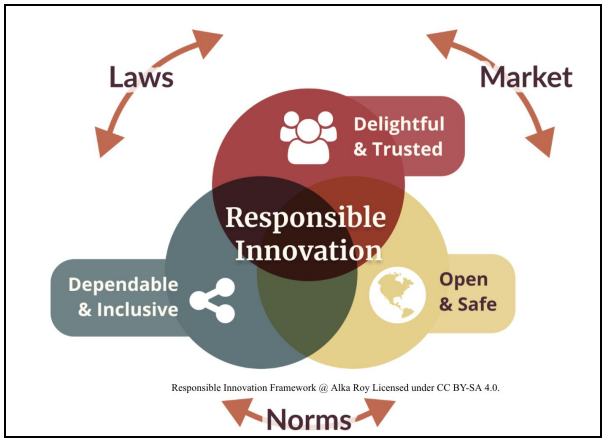

Figure 2: Responsible Innovation Framework

Frameworks mean different things to software developers, theorists, or those probing patterns of behavior. The framework presented in this paper is collective and interdependent. It encompasses many other rubrics and decision frameworks that need to be further probed.

Trust, delight, dependability, inclusion, openness, or safety can be interpreted and applied differently for and by different stakeholders while incorporating business interests and good design and engineering principles. This framework is a start to include the considerations that impact our society as "essentials," instead of "nice to have."

The specific application of the framework would vary depending on stakeholders, use case, level of risks, and impact. One workshop used this framework as a reference to existing design thinking and iterative agile methodologies. The questions that this framework raised: 1) What is missing in current considerations? 2) Is there room to push cultural or habitual boundaries to create opportunities for more inclusion, delight, trust, etc? 3) Is there a multi-variable approach instead of single metric decision-making? 4) What happens to the design variables and decision-making criteria when external pressures are viewed as influences and the idea of stakeholders are expanded?

The following table shares considerations for applying the framework and illustrates how themes from Trusted AI and Ethical guidelines can be mapped to the framework.

The Responsible Innovation Framework		**Applying the Framework**	**Mapping to Ethical & Trustworthy AI Guidelines**
Overall	**Highlights: Innovation, Reframing Interdependence, Integration into existing organizations, Accessible Language, and Visual Representation**	**Apply at multiple decision gates: From ideation to feature prioritization and description, testing & reassessment, etc.**	**Highlights: AI-specific, List, Governance, Introduces language of ethics and people-centered, Builds New organizations**
Value-Sets	**Delightful & Trustworthy**	Are we giving our users a viable choice (trust) to get a personalized experience (delight) as quickly as they might want (delight) without storing, extracting, using, or selling their data and extracting consent in a way that makes them vulnerable (trust)?	Delight is often explicitly not mentioned in Ethics considerations. Trust is covered by FAT (Fairness, Accountability, Transparency and the general ethical framework.
	Dependable & Inclusive	If the product provides basic access to services, are there equivalent alternatives for the users or groups that are not included? Can potential users go unserved in an effort to make the system robust for a few?	Dependability maps to Robustness and Accountability. Inclusion maps to equity and fairness. Interdependence not explicitly explored.
	Open & Safe	Given the particular application, goals, and circumstance, how open can the innovation be while being as safe as the public or its users need it to be?	Maps to Transparency, Explainability, Privacy & Security. Interdependence is not explicitly explored.
Stakeholders	**People**	Who is it for? Why do they need it? Who is included or excluded? Who might be exploited or harmed? Will they like us if they found out? Would we like us if everyone found out? How does it impact you and us as people?	Human-Centered AI, Well Being
	Environment	Can we not only make it using less stuff (economically) so that it would useless stuff? Can we make it so it has either a longer shelf-life or creates less waste when disposed of? Could we possibly make it so that we would generate and recreate what it needs?	The planet is occasionally mentioned, primarily as sustainability
	Things	Do people know and comprehend that they are interacting with a set of hardware and software systems developed and operated by a team of people? The communication between systems and machines is different as it is between humans or humans and machines.	Primarily Addresses Products or Use cases for Robotics/AI/ML/Data/Software
Influences	**Market**	Allows for Sustainability. But should be one of the metrics with checks and balances. Business accountability	Seen as adversarial or primary transactional
	Legal	Many current laws can be applied to people. The challenge is that the legal system itself may need innovation.	Governance, Accountability, Explainability & Transparency. Regulatory & Policy
	Norms	Self-regulation and external accountability	Common Standards, internal Compliance, Assessment, Governance

Table 1: Applying the RI Framework (Includes a Mapping to Common Themes in Ethical & Trustworthy AI Guidelines)

5. The Collective Framework: An Example

The Responsible Innovation Framework is not just for high stakes use cases with large social implications, like security and privacy for children's use of the internet, bias mitigation, surveillance technology, or harm in the hands of bad actors that often ends up in the news. The reality of technology making is that many small decisions by well-meaning people and teams can accumulate into something harmful and problematic and lead to a loss of trust. It is important to build the muscle and habit of thinking about social impacts.

Consider this simple real-life example: A product manager is attending a vendor meeting in an unfamiliar part of a congested city. She gets an urgent message that her child was injured at school. She rushes out and opens up a map app on her phone known to find the fastest route in high traffic areas.

On the way, she reaches a busy intersection. From scanning the directions before her trip, she knows she is supposed to turn but is unsure which lane to take because the audio is muted. Cars are lining up next to her. She glances at the screen, hoping to spot the left or right arrow. But a pop-up ad for a nearby car dealership is covering half her screen and the directions, and she misses her turn. How did so many smart people design this terrible feature?

Let's start with influences. What is the design or accepted understanding (Norms) for keeping maps ad-free or what about legal requirements to keep the eyes on the road (Legal) or driving safely? And then there is the influence that led to wanting to monetize the mapping app or keep it free because of the fear of backlash and lack of use (Market). The stakeholders would be the *people* using the app while driving or riding the car, the daughter waiting for her mother, the school staff, the other people driving on the road next to the car where someone is using the app, etc, the things (app, phone, car) and the environment (the level of congestion or urgency or likelihood of needing whatever the ad is selling).

Considering that safety and dependability rank high for maps, the ad placement is difficult to justify. But even if the desire to monetize or delight by offering a coupon or deal could be met, when the ad is shown (not at an intersection or when the user is driving fast) and considering what the ad is shown for (not car dealerships that have a low likelihood of being motivated by ads vs ice-cream or coffee) and where the ad is placed on the screen would increase both delight and trust. Finally, the user could be given a choice to "opt-in" or set a preference for ads for navigation and even have a pop up before the start of the trip to navigate the competing interests.

6. Future Work

The Responsible Innovation Framework intends to serve as a visual reminder for multiple and expanded considerations of stakeholders, influences, and values. A group of engineering leaders and directors from US companies and a South African fintech researcher have applied the framework to Xiaoice, the teenager-girl-like AI chatbot case study as well as other products. The biggest impact so far appears to be the act of assessing a product with the framework itself because social variables become central and essential.

More inquiry, data, and work are needed, including 1) further development of theoretical concepts behind the framework components like "technology as a stakeholder" 2) creating a repository of examples to illustrate stages and degrees of responsible innovation including with AI and another emerging tech, 3) exploring strategies (like incentives) and impact of applying and adopting the framework, and 4) considerations for specific domains and industries, to make it easier to shift to responsible innovation.

7. Conclusion

The only reason the "responsible" qualifier is added to this framework is the need to balance values and stakeholders is lacking in the current practice of technology innovation and application. Otherwise, this *is* a framework for innovation that can lead to greater trust. In the current ecosystem, we collectively seem to need a reminder that success and responsibility can coexist. That sometimes slowing down to consider the implications speeds up adoption and avoids the unnecessary human cost and environmental impact. Especially when innovation is transforming every aspect of our individual and social interaction at work, home, public systems, education, healthcare. The reason the framework is not relegated to a particular discipline or area is that technology is not relegated or limited to a particular area, and its portability needs to be mirrored in any development framework.

Every decision, every feature, every new technology evolution is an opportunity to change the narrative of innovation. Instead of labeling and classifying people or companies as "responsible" or "irresponsible", "trustworthy" or "honest," this framework intends to shift the evaluation of each decision with the opportunity to do things "responsibly" or "ethically." The technology industry has the habit of iterating and this framework serves to include the goals of a more delightful, trustworthy, dependable, inclusive, open, and safer future, one feature and step at a time.

Ultimately, the relevance and shift to responsible innovation depend on many internal and external factors, incentives or lack of incentives through regulations, actionable auditing, or successful examples that illustrate and normalize social considerations into current processes. As with any framework, the effectiveness depends on how someone is interpreting and applying it. Managing conflicting

values and shifting to greater responsibility is a messy process without clear answers. So is innovation. It is a big, difficult, and exciting responsibility. That is why the proposed future work of iterating on the framework, creating a repository of examples and case studies, and laying out key considerations and challenges is extremely important if our goal is to build greater trust.

9. Acknowledgments

Like everything, this is a work in progress, and I take full responsibility for any shortcomings in my analysis, thoughts, or ideas. This framework is a synthesis of ideas inspired by many others. It is a continuum and offshoot of the works cited here and countless other conversations that have influenced me along the way. I am grateful to all those influences, my colleagues, and the anonymous reviewers for their thoughtful and critical suggestions. I'm especially thankful to John Holleman, who helped take my early visual representations and greatly improved them.

9. References

[1] Doreen Bogdan-Martin, "Facts and figures 2019. Measuring Digital Development Report", ITU, Telecommunication Development Bureau, 2019.

[2] Julia Carrie Wong, "The Cambridge Analytica scandal changed the world — but it didn't change Facebook", The Guardian, March 18, 2019.

[3] Cade Metz, "The Week in Tech: YouTube Fined $170 Million Over Child Privacy Violations", The New York Times, September 6, 2019.

[4] Kevin Roose, Sheera Frenkel, and Nicole Perlroth, "Tech Giants Prepared for 2016-Style Meddling. But the Threat Has Changed", The New York Times, March 29, 2020.

[5] "Ethically Aligned Design (EAD1e): A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems.", The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, IEEE SA, 25 March 2019.

[6] Fjeld, Jessica and Achten, Nele and Hilligoss, Hannah and Nagy, Adam and Srikumar, Madhulika, "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-Based Approaches to Principles for AI", Berkman Klein Center Research Publication No. 2020-1, January 15, 2020.

[7] Edwards-Schachter, Mónica, Wallace, Matthew, 'Shaken, but not stirred': Sixty years of defining social innovation, Technological Forecasting and Social Change 119 pp. 64-79, March 2017.

[8] Kristen Pue, Christian Vandergeest, and Dan Breznitz, "Toward a Theory of Social Innovation" Innovation Policy Lab White Paper 2016-01, December 1, 2015.

[9] Schomberg, René von (2012) 'Prospects for Technology Assessment in a framework of responsible research and innovation' in: Technikfolgen abschätzen lehren: Bildungspotenziale transdisziplinärer Methode, Wiesbaden: Springer VS, pp. 39-61.

[10] Floridi L, Cowls J. A Unified Framework of Five Principles for AI in Society. Harvard Data Science Review [Internet]. 2019 Jul 1;1(1).

[11] Turing, A. M. "Computing Machinery and Intelligence." Mind, vol. 59, no. 236, 1950, pp. 433–460.

[12] Banaji, Mahzarin R. author. Blindspot: Hidden Biases of Good People. New York: Delacorte Press, 2013.

[13]Hagendorff, Thilo. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. Minds and Machines. 10.1007/s11023-020-09517-8.

[14] H. Van Lente, Promising Technology - Dynamics of Expectations in Technological Developments (Thesis), Twente University, Enschede, 1993.

[15] H. Van Lente, A. Rip, Expectations in Technological Developments: C. Disco and B. van der Meulen (Eds.), Getting New Technologies Together Expectations in Technological Developments: An Example of Prospective Structures to be Filled in by Agency, Walter de Gruyter, Berlin-New York, 1998.

[16]Taddeo, Mariarosaria and Luciano Floridi, 2018, "How AI Can Be a Force for Good", Science, 361(6404): 751–752.

[17] Schomberg, René von (2013). "A Vision of Responsible Research and Innovation" Responsible Innovation. Responsible Innovation: Managing the Responsible Emergence of Science and Innovation in Society. John Wiley & Sons, pp. 51–74.

[18] Jack Stilgoe, Richard Owen, Phil Macnaghten. "Developing a Framework for Responsible Innovation", Research Policy, Elsevier, Vol.42, Is 9, November 2013, pp. 1568-1580.

[19] Jordan Novet, "Microsoft spins off 'virtual teenager' chatbot for Chinese users", CNBC, July 13, 2020.

[20]Wertheim EH, Paxton SJ (2011) Body image development in adolescent girls. In: Cash T, Smolak L (Eds.), Body image: A handbook of science, practices, and prevention. (2nd ed), The Guilford Press, New York, NY, USA, pp. 76-84.

[21] Cole, Raymond (2012). "Transitioning from green to regenerative design". Building Research & Information. 40: 39–53.

[22] Felzmann H, Villaronga EF, Lutz C, Tamò-Larrieux A. Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. Big Data & Society. January 2019.

[23] Microsoft opens AI framework to other firms, China Daily, August 22, 2019

[24] Microsoft's politically correct chatbot is even worse than its racist one, Quartz, July 13, 2018

[25] Lessig, Lawrence "The New Chicago School". The Journal of Legal Studies. 27 (S2), June 1, 1998, pp. 661–691.