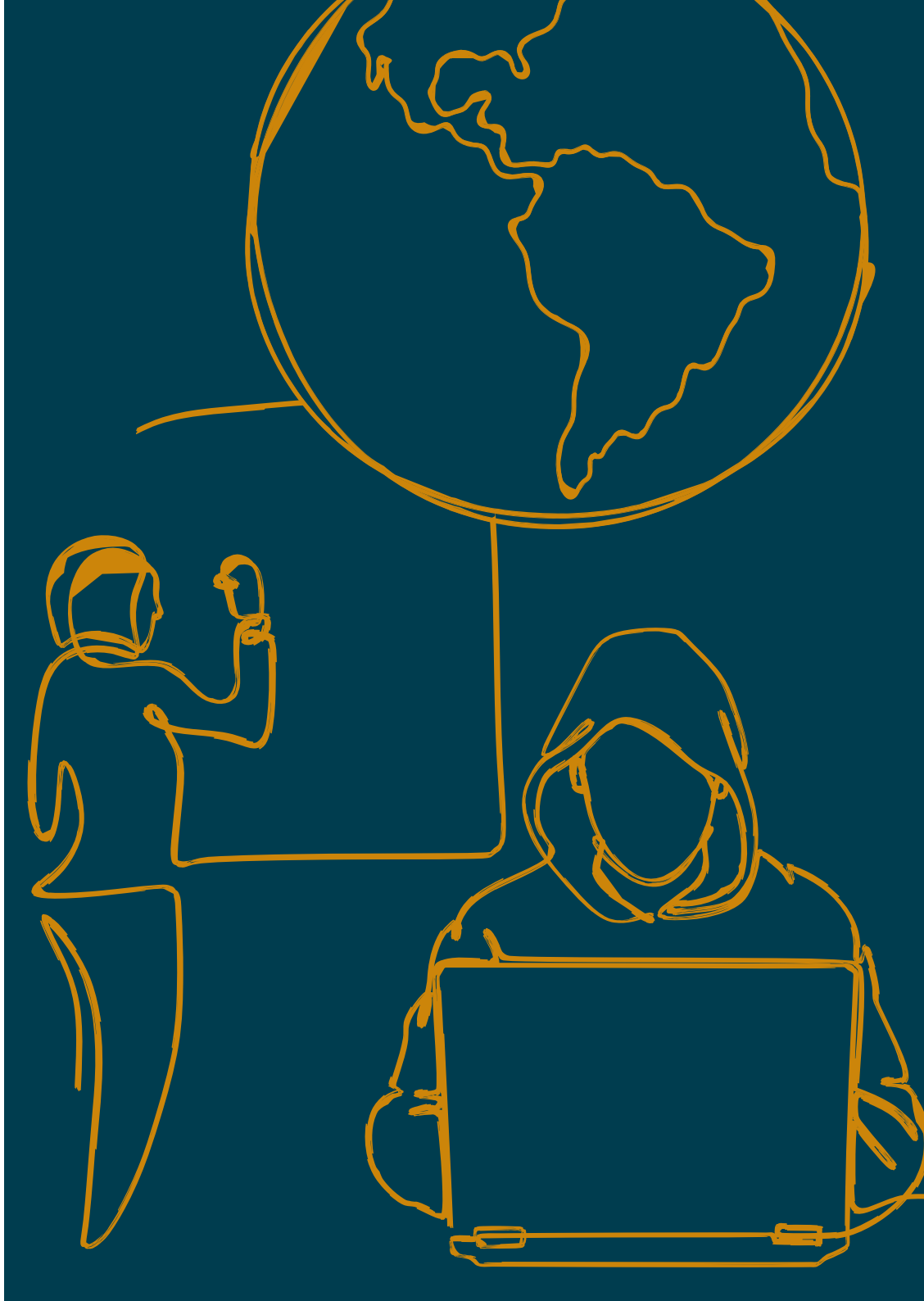
/thoughtworks

Responsible tech playbook

A guide to the tools and practices that help businesses make better technology decisions



August 2021

© 2021 Thoughtworks

The content of this playbook that is created by Thoughtworks is subject to the Creative Commons Attribution Share Alike 4.0 International license (**https://creativecommons.org/licenses/by-sa/4.0/** / **https://creativecommons.org/licenses/by-sa/4.0/legalcode**).

There are tools and images within the playbook that are created by and owned by the respective tool creators, and such material is included with their permission. Any tools and related images are subject to the license(s) of the respective tool creators. If you wish to use any of the tools and images in the Playbook please contact those tool creators/rights holders directly to check the compatibility of the licence(s) with your planned use.

Thoughtworks does not endorse or recommend any of the tools, methods or frameworks found in this playbook; inclusion should not be construed as an endorsement or recommendation.

/thoughtworks

© 2021 Thoughtworks

© 2021 Thoughtworks

Contents

Foreword 4
Our mission 5
What is Responsible tech? 6
Responsible tech: why it matters 8
Responsible tech tools and practices 10
Responsible tech in the development lifecycle 11
Choosing the right tools 12
Methods and tools index 13
Tools to get started 14
Responsible technology tools in-depth 15
- Agile Threat Modeling 16
- Consequence Scanning 18
- Data Ethics Canvas 20
- Ethical Explorer 22
- Ethical OS 24
- Failure Modes & Effects Analysis 26
- Flourishing Business Model Canvas 28
- InterpretML 30
- Materiality Matrix Assessment 32
- Responsible Strategy 34
- Tarot Cards of Tech 36
- Unintended Consequences 38

Continuing the Responsible tech journey 40
Promising approaches 41
Other emerging approaches to explore 46
Further reading 47
Acknowledgments 47

Foreword

The reach of technology is extending into more and more sensitive and complex arenas, from credit decisions and medical diagnoses to sentencing — to name just a few. It impacts everyday interactions with ourselves, friends and family; it affects our lives as employees, customers and citizens. We must, as technologists, actively take responsibility for these impacts, and the unintended consequences of our work — amplifying the positive and mitigating the negative.

The tools and techniques in this Responsible tech playbook help teams identify strategies to be more inclusive, aware of bias, transparent and to mitigate negative unintended consequences. Using these approaches helps technologists examine their product and technology choices from multiple perspectives, increasing the likelihood that harmful or unintended consequences will be uncovered.

While our industry doesn't have an analogue to the Hippocratic oath or professional licensing standards, as found in other professions, we still need to consider what responsibilities we have to those who use the technology that we build, as well as those who are invisible during the design process. This playbook is a way to start building Responsible tech from the beginning.

Rebecca Parsons
Thoughtworks CTO

© 2021 Thoughtworks

Our mission

Helping organizations navigate the ethical risks emerging in an increasingly digital world

Businesses are increasingly aware the technology they create can have massive ethical (and associated reputational and financial) consequences. When developing products, more organizations are trying to tread carefully — but technology is a vast, complex and shifting field, and when it comes to measuring the potential impact of what they're building, people often aren't sure where to start.

This book highlights some of the leading tools that any organization or team can realistically adopt. The tools can help analyse technology and development processes for potential ethical challenges, and start on the path to a more responsible technology practice. The emphasis is on practical, participatory techniques that anyone — not just technology specialists — can adopt and benefit from, and use to increase their understanding of how technology interacts with society.

The playbook is like a cookbook: start using the recipes but also get inspiration from others, find out what suits your context and develop your own skills. Start small, use it frequently and build your culture and capabilities — as an individual, a team, and an organization. It's like a muscle you build with regular exercise.

We would love to hear from you about your experiences, other tools, tips on how you use them, and the impact they made. The goal is to further support and catalyze the successes made in this field.

Alexander Steinhart
Editor & on behalf of the playbook's stewards

© 2021 Thoughtworks

What is Responsible tech?

Aligning technology and business behavior with society's and individual's interests.

Responsible tech is a way of working that aligns technology and business behavior with society's and individual's interests. It explores and actively considers the values, unintended consequences and negative impacts of tech, and actively manages, mitigates and reduces risk and harm. We consider notions of ethics, individual and human flourishing, social structures, inclusivity and equity, civil liberties and democracy. We are striving to reach a state where technology doesn't exploit us — it supports us.

This is driven by digital technologies which are becoming ever closer to ubiquity. They move at the core of our individual use and of our societies. As people rely more on technology for daily activities, they are also more subject to unintended — or downright hostile – consequences.

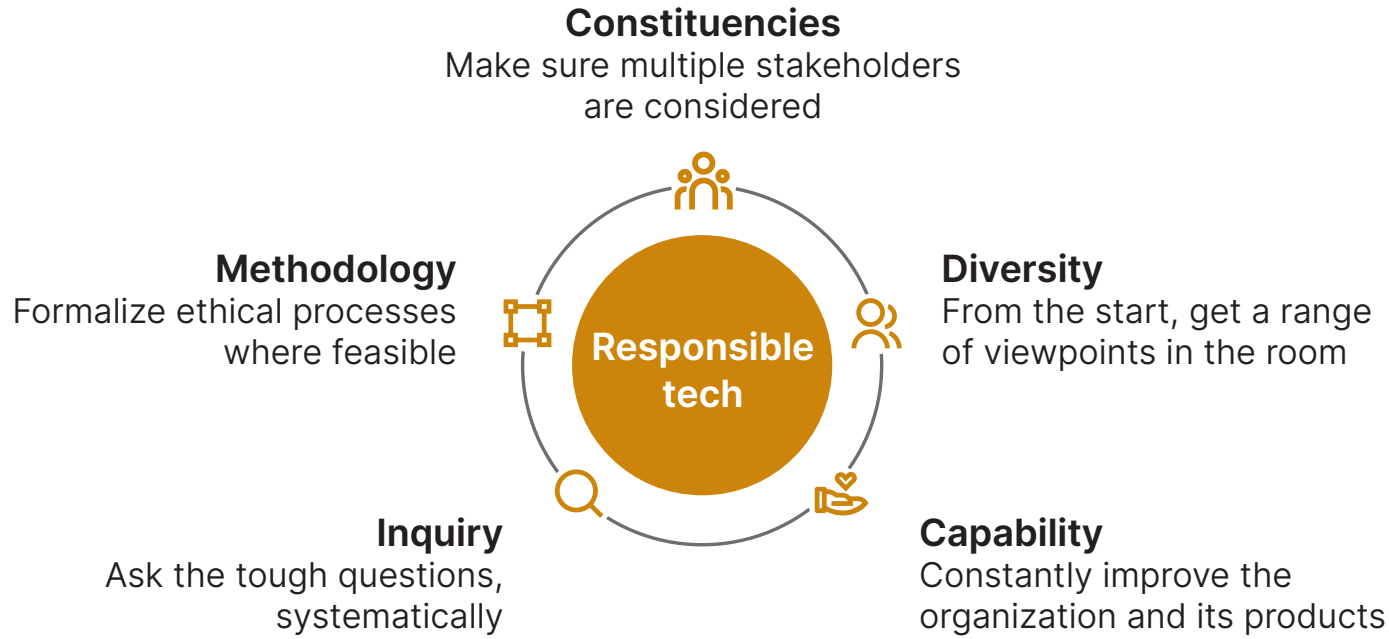
The combined effect of greater automation — action with real-world consequences without human intervention — and increasing complexity of technology systems leads us to less predictability and more significant impact. Therefore, the need for such mature risk management and security practices has never been higher.

At the same time, "tech" and "society" are becoming increasingly inextricably intertwined, with technology affecting human behavior and human behavior affecting technology. Therefore, the need for such value-explicit practices on how we create futures that our communities and we desire is high.

© 2021 Thoughtworks

Operationalizing Responsible tech

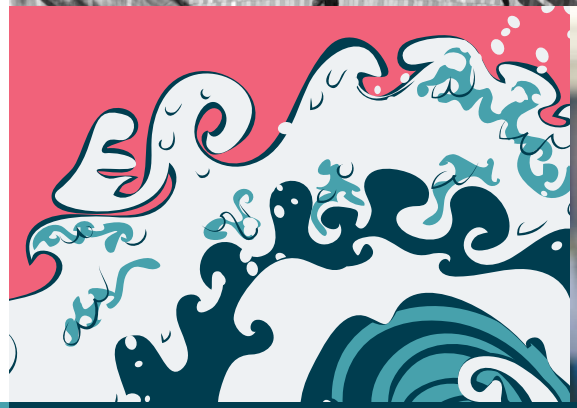
When you seek to further embed and operationalize Responsible tech — to keep the ability to innovate with greater confidence, and foster trust with stakeholders, consumers and the broader public — these five fields will help you to get started: inquiry, methodology, constituencies, diversity, and capability.



The Responsible tech tools in this playbook will help you with methodology and inquiry, formalize processes, and systematically ask the tough questions. However, you still need to ensure that multiple stakeholders (constituencies) and a range of viewpoints (diversity), including ethics experts, are considered. Lastly, you need to allocate resources to build and improve this culture and capabilities continuously.

© 2021 Thoughtworks

Responsible tech: why it matters



Demands (or impact) and benefits

- **Ninety-seven percent** of consumers expect brands to use technology ethically while innovating (**source**)
- **Seventy percent** of consumers say improved ethics make them more likely to buy from a brand (**source**)

Challenges and consequences

- Only **nineteen percent** of people believe tech products and services are designed with their best interests in mind (**source**)
- **Eighty-six percent** of consumers are likely to sever ties with businesses that use data unethically (**source**)

© 2021 Thoughtworks

The consequences of **not** incorporating responsible and ethical thinking can be devastating to our identity, health, business, economics, political systems and our social cohesion. As technologists, the tools we apply can help prevent these drastic consequences.

Explore the broader Responsible tech context in our **Social impact report: Tech at the core of society**.

© 2021 Thoughtworks

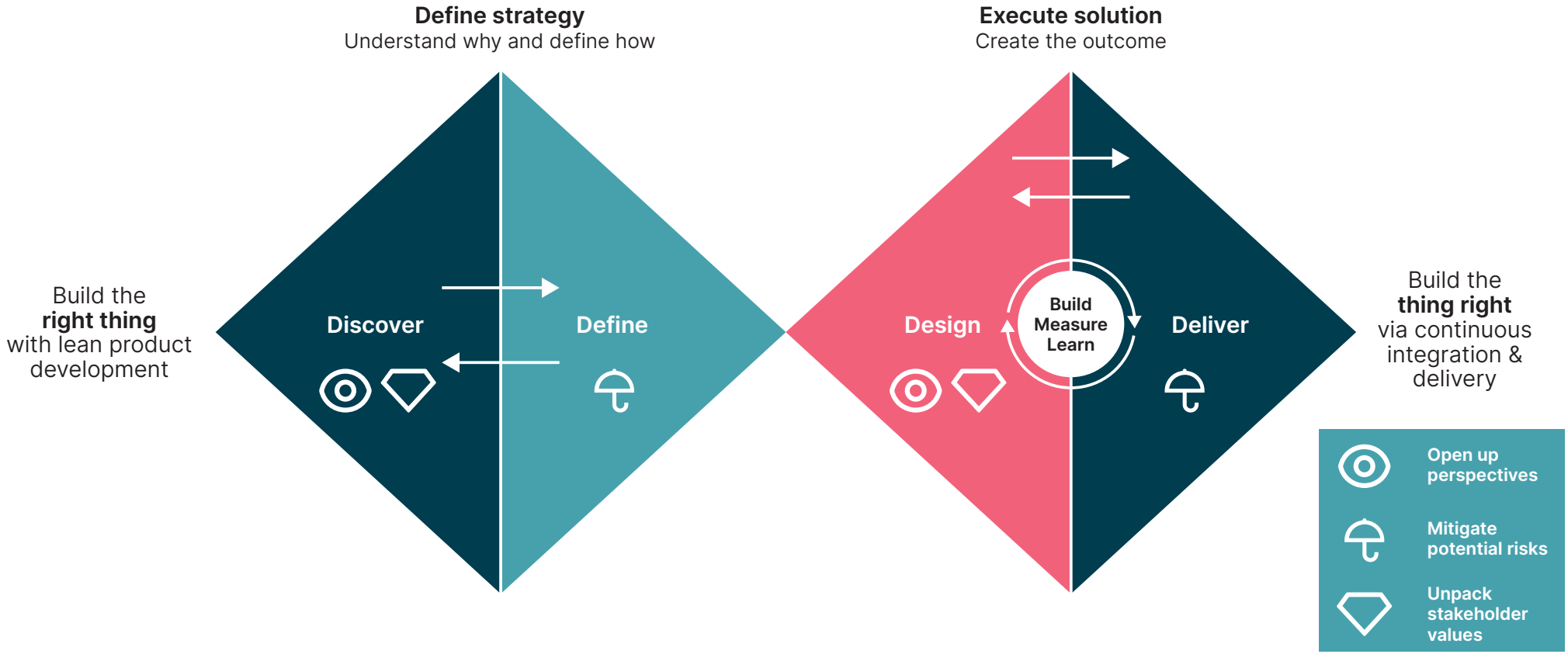
Responsible tech tools and practices

© 2021 Thoughtworks

Responsible tech in the development lifecycle

You can integrate responsible tech at all stages of the product development cycle. During the divergent phases, tools that open up perspectives and unpack stakeholder value are especially suitable. During the convergent phases, mitigating tools magnify.

See how to further operationalize Responsible tech on page seven.



© 2021 Thoughtworks

Choosing the right tools

Use this key to help you decide which tools are best suited to your team's situation and goals

Open up perspectives

Solicit different points of view to think through a wider range of potential consequences and outcomes.

Mitigate potential risks

Identify and address ethical challenges and vulnerabilities before they become bigger problems.

Unpack stakeholder values

Ensure the technology is designed to meet the needs and support the values of those it is intended to serve.

© 2021 Thoughtworks

Methods and tools index

In alphabetical order

Agile Threat Modeling

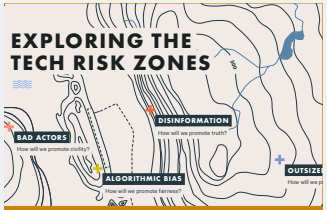
page 16

Consequence Scanning

page 18

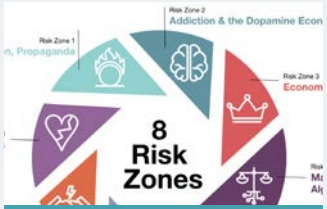
Data Ethics Canvas

page 20



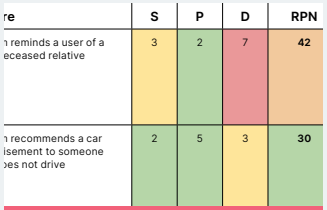
Ethical Explorer

page 22



Ethical OS

page 24



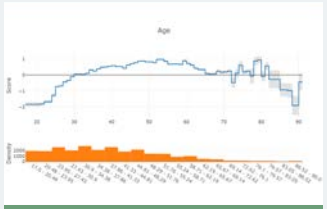
re	S	P	D	RPN
n reminds a user of a eceased relative	3	2	7	42
n recommends a car isement to someone es not drive	2	5	3	30

Failure Modes & Effects Analysis

page 26

Flourishing Business Model Canvas

page 28

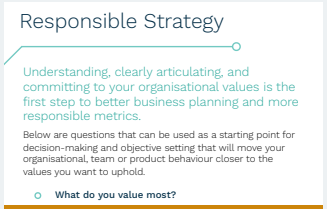


InterpretML

page 30

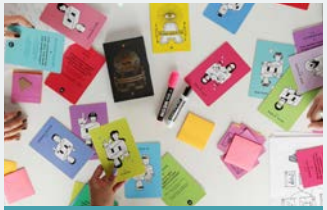
Materiality Matrix Assessment

page 32



Responsible Strategy

page 34



Tarot Cards of Tech

page 36

Unintended Consequences

page 38

For additional approaches and methods, see pages 41–47.

© 2021 Thoughtworks

Tools to get started

Unsure where to start? Begin to explore by trying one of these three tools that we found useful to take the first steps. Any of them is perfectly suited to get you started with the topic of Responsible tech generally and are suited to be used in an inception or start of a product.

Ethical Explorer

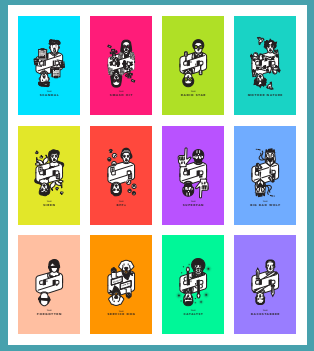
The Ethical Explorer offers prompts (mostly based on the Ethical OS) and guidance which can be used in individual settings but also with the whole team and stakeholders. It helps you to get started thinking about what ethics is and the various aspects of risk that might be relevant to consider with your product.

Learn more >

Consequence Scanning

Consequence Scanning enables teams to not only address the right questions but also to empower them to take action to make the relevant changes. The structure will be very similar and easily adaptable to any team that runs agile ceremonies such as team retrospectives.

Learn more >



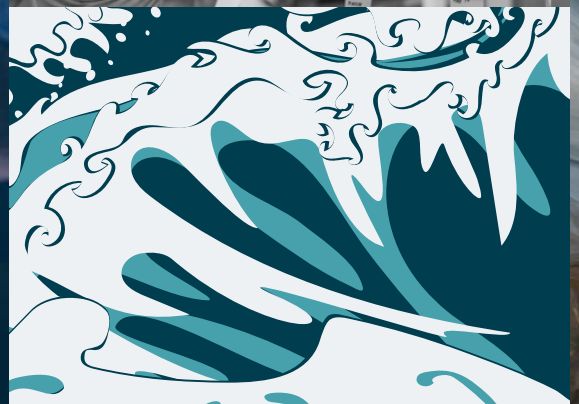
Tarot Cards of Tech

Tarot Cards of Tech is a beautifully designed set of prompts that open up perspectives and help you to take perspectives. They pair well with creating HMW statements in ideation, and could provide great guiding questions to add to your processes.

Learn more >

© 2021 Thoughtworks

Responsible technology tools in-depth



© 2021 Thoughtworks

Agile Threat Modeling

By Thoughtworks

What is it?

Threat modeling is a risk based approach to designing secure software based on identifying threats and mitigations. Rather than creating and maintaining an exhaustive "threat model" we encourage teams to do threat modeling "little and often".

When should we use it?

Get your teams to all be in the habit of threat modeling every iteration. Bring the software delivery team together to brainstorm genuine security threats before they materialize.

Who should be involved?

Involve the whole delivery team in each session, which is to say both technical and non-technical roles. For example, include business analysts and product managers to raise awareness and get the right risk perspectives. It can be the perfect opportunity to get insight from someone from the security team, but not required.

< Return to methods list

Mitigate potential risks

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

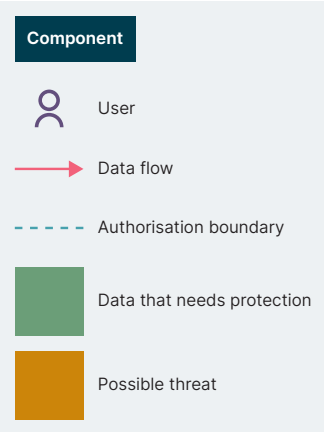
- Do not bite off too much at once! Pick an achievable scope
- Use the STRIDE cue cards to help brainstorm, and extend the cards with other dimensions (e.g. cards highlighted in this playbook)
- Make sure the actions go straight into the backlog

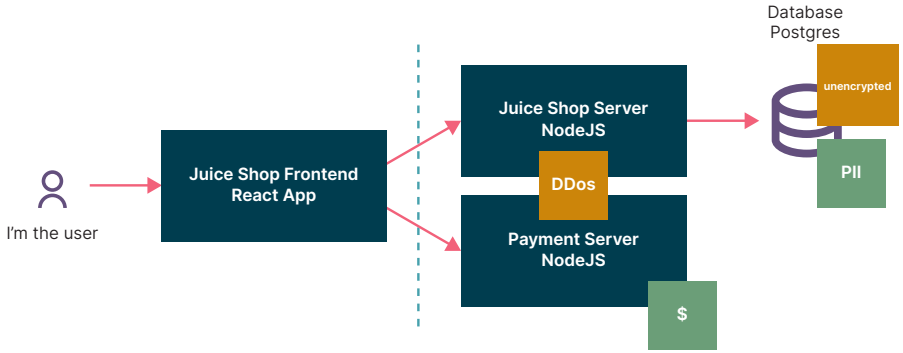
Explore further:

Full guide to thread modeling (no download required)
[Article by Jim Gumbley; Getting started guide for the workshop.]

STRIDE cue cards available in Thoughtworks UK, Germany and India offices

Agile Threat Modeling (continued)





How does it work?

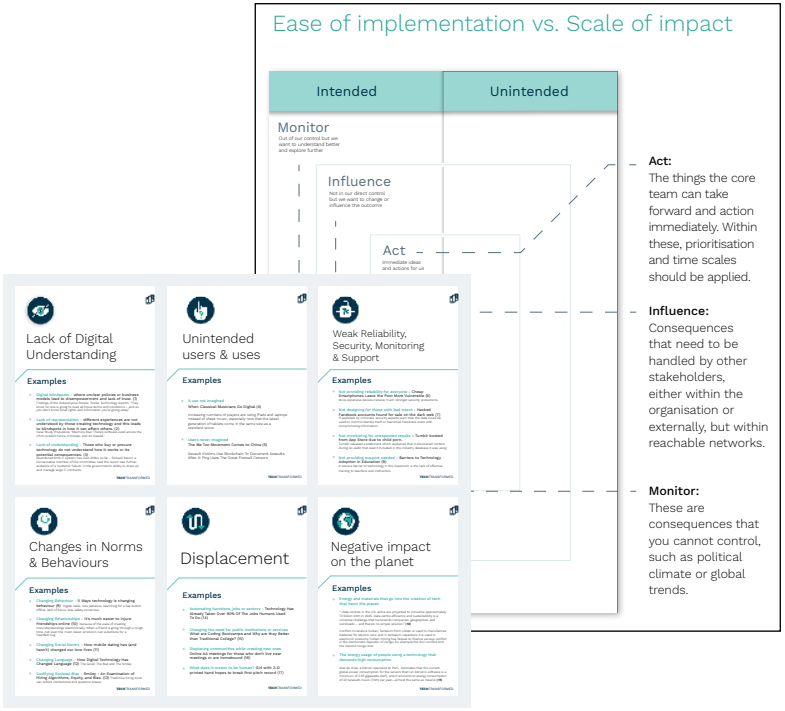
1. Gather delivery team around a whiteboard for frequent, short threat modeling sessions asking:
   - What are we building?
   - What can go wrong?
   - What are we going do about it?
2. Explain and explore by beginning to draw a 'lo-fi' technical diagram and include:
   - Relevant components and users
   - How data flows in different directions
   - Label networks and boundaries
   - Assets important to business value
3. Brainstorm threats using the STRIDE cards
   - Coming up with ways to attack, break or frustrate a particular bit of software is threat modeling at its essence
   - Capture the threats that are identified either on a board, stickies or online collaboration tool (ie. MURAL)
4. Prioritize and fix by sharing knowledge that is useful for prioritization
   - Team should dot vote about three of the riskiest threats
   - Identify which ones are the riskiest and document
   - Take action by capturing steps to mitigate in your backlog
5. Wrap up and close
   - Assign actions to members of the team
   - Decide when the fixes could be implemented
   - Try the method again once the initial fixes have been delivered

< Return to methods list

© 2021 Thoughtworks

Consequence Scanning

By Doteveryone (now Open Data Institute)



What is it?

This kit provides workshop activities and materials that help teams consider the intended and unintended consequences associated with a technology or a service. It's a tool that can be used for all teams, across any domain. Participants should be as cross functional as possible: from technology, design, product, business, as well as users themselves.

It also provides a framework to prioritize issues teams might uncover as a result of the workshop.

In a structured session, the following questions are covered:

- What are the intended and unintended consequences of this product or feature?
- What are the positive consequences we want to focus on?
- What are the consequences we want to mitigate?

When should we use it?

It's designed to fit into the iterative agile process, so is suitable for multiple stages including the initial conception, roadmap planning and when new features are introduced.

- It's easy to use, with very comprehensive instructions
- Try to allow 1.5–2 hours for the session, otherwise it can feel rushed

Plug it into: vision, ideation, implementation, roadmap, or epic-level and project wrap-up (for sales follow-up) — make it a habit. It can become part of a team's normal activity, like a retro. The tool is reflective and well-balanced, action orientated, and team-empowering.

Who should be involved?

Core team, user advocates, collaborators such as technology or business specialists, senior product sponsors, collaborating senior stakeholders.

< Return to methods list

Mitigate potential risks

Open up perspectives

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

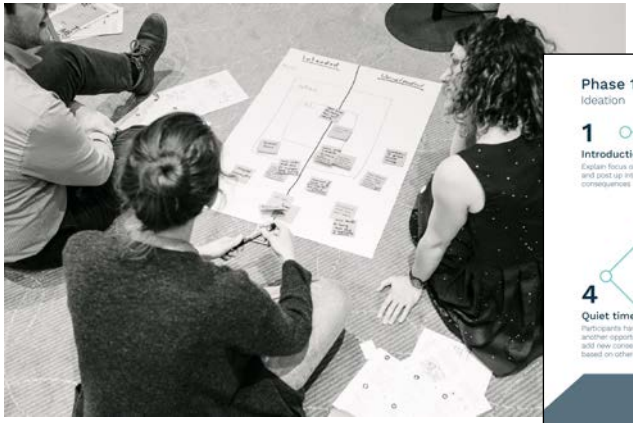
The prompts can be a little abstract and hard to connect to. Take time to collect the most relevant prompts so you can help participants relate. You can also use prompts from other tools such as Tarot Cards of Tech, Ethical Explorer and so on. However, it's an amazing Agile tool that fits well together with ceremonies such as retros and reviews.

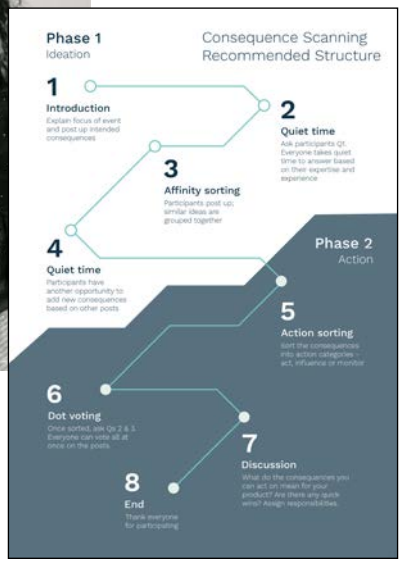
Explore further:

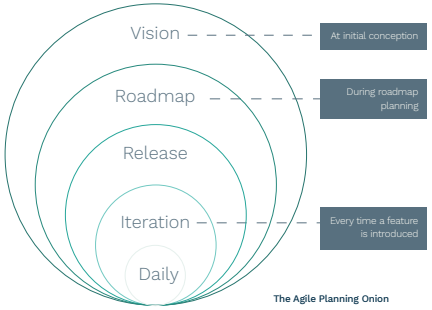
https://www.doteveryone.org.uk/project/consequence-scanning/

Originally produced by **Doteveryone** (now maintained by the **Open Data Institute**) Attribution - Noncommercial ShareAlike 4.0 International (CC BY-NC-SA 4.0)

Consequence Scanning (continued)







How does it work?

The method has two phases:

Phase 1, Ideation: Different questions and prompts to help uncover intended and unintended consequences (introduction, quiet time, affinity sorting and collecting).

Phase 2, Action: Sort outcomes into categories to help define an action plan (action sorting, dot voting, discussion and end).

Consider:

- If doing remotely, split into smaller groups and use a shared wall or whiteboard for each group. Choose a tool that accommodates real time collaboration
- Get different perspectives about your focus into the room. For example: people who have strong opposing views, or have experienced this 'feature', or who have never encountered it before. This helps drive out more richness in the analysis

Also include:

- Your company's vision, mission and values
- Other tools or documents that support reflection and decisions

< Return to methods list

© 2021 Thoughtworks

Data Ethics Canvas

By Open Data Institute

What is it?

A collection of questions, which aim to prompt discussions so that teams consider how data is to be collected, shared or used.

Aims:

- Find a balance between data 'fearing' and data 'hoarding'
- Consider impacts on people
- Encourage trust and be transparent

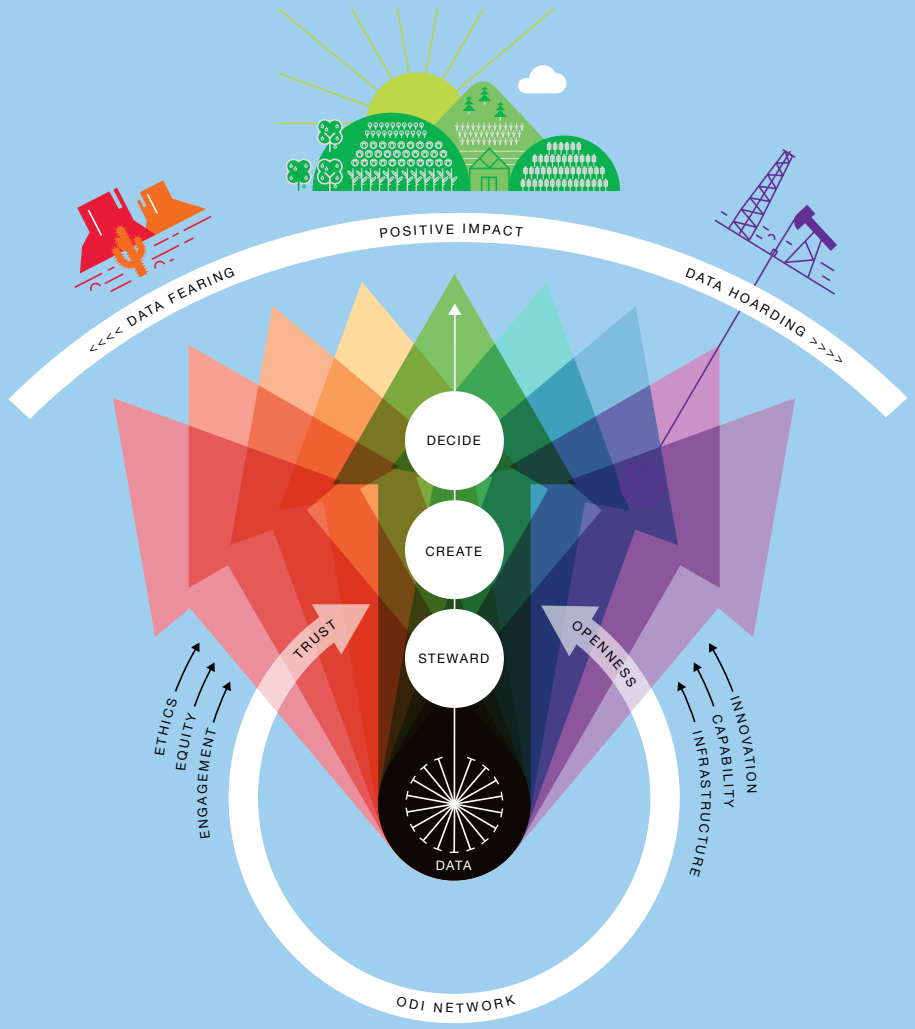
Outcome:
Produce a prioritized and assigned list of actions.

Why use it?

Helps to identify potential ethical issues in data usage, and what mitigating actions could be taken.

When? Ideally at the start of a new project, but still helpful later
With whom? Activity is best for a diverse group.

< Return to methods list



Mitigate potential risks

Open up perspectives

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

- Document your notes and actions
- Complete the canvas in a group with a range of people
- Be aware of how power dynamics in the group might impact the exercise
- Revisit the answers and actions regularly

Explore further:

https://theodi.org/wp-content/uploads/2019/07/ODI-Data-Ethics-Canvas-2019-05.pdf

Produced by the **Open Data Institute**. Licensed under a Creative Commons Attribution-ShareAlike 4.0 UK International Licence.

Data Ethics Canvas (continued)

How does it work?

You can start anywhere on the canvas and complete the sections in any order.

1. For each section, take time to consider your answers. It helps to note them on sticky notes and stick them to the sections as you go.
2. Where you can, share and discuss your responses with others.
3. When you have completed the canvas, outline next steps and ensure there is a person responsible for each action.
4. Share notes about the discussion more widely if you can.

< Return to methods list

© 2021 Thoughtworks

Ethical Explorer

By Omidyar Network

What is it?

A set of cards that help you explore the different risk zones that might arise from your software product. Using the metaphor of an explorer and risk zones as the terrain, imagine the current landscape that your own product or service resides in and scout the potential dangers that could lie ahead. It is similar to Ethical OS, so both methods should be tried and compared.

- Using metaphors and prompts to help you identify areas where your product has risks and the actions you can take to mitigate them
- Can be used in any domain

Why use it?

Encourages thinking beyond the normal project risks by providing questions that will open up discussion about identifying what those dangers could be and what to do thereafter.

When? Could be used at any stage of product delivery, however the earlier the risks are identified the easier they will be to mitigate.

With whom? Should include senior stakeholders during earlier stages of a project, while it can also be used during any stage with a cross functional team (two or more people).

< Return to methods list

Mitigate potential risks

Open up perspectives

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

The cards are a good iteration to the EthicalOS, giving more guidance on how to apply it. Further, the imagery and metaphors used are well suited and encourage people to explore the topic. Various examples on how to apply the cards, whether tackling alone or as group activities, give a good rounding.

Tool produced by **Omidyar Network** and **Artefact**.

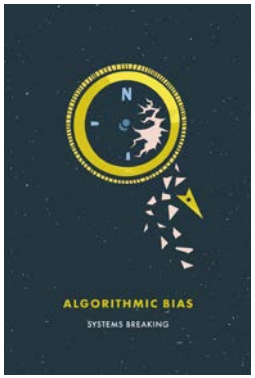
Explore further:

https://ethicalexplorer.org/download/

Ethical Explorer (continued)

How does it work?

1. Start with a warm up exercise by having each participant go through the cards independently, and then share with the group an example of a product that has potential risk that could go under one or more of the risk zones.
2. Continue by getting the participants to think about which of the risk zones the product or service could fall under. Have them pick up to three of the zones.
3. Next, pick the top zone or zones chosen by the group, and begin to brainstorm using the prompts on the back of the cards.
4. Discuss the results and ask the group 'How could this affect the product/service over time? Which is most detrimental?'
5. Decide how to prioritize issues by dot voting: You might ask 'Which has the most detrimental effects? Which is easier to change?'
6. Brainstorm on ways to mitigate those issues as a group and steps toward those actions.



< Return to methods list

© 2021 Thoughtworks

Ethical OS

By Omidyar Network's Tech and Society Solutions

What is it?

EthicalOS is a toolkit for business, product and tech to shape strategy, process and values of a company, its products and services. It is similar to Ethical Explorer, so both methods should be tried and compared.

It includes:

- A checklist of eight risk zones
- 14 future scenarios to make risk zones more tangible
- Instructions on how to use risk zones and scenarios in a workshop context
- Strategic measures towards ethical tech
- Custom use cases

Why use it?

EthicalOS is a strategy toolkit that works best in a discovery or advisory stage with major client stakeholders involved.

However it could also be used:

- To prepare for a project in any phase (even sales) to articulate concerns and create tangible scenarios for undesirable futures
- To access existing solutions and understand risk areas to pay attention to

< Return to methods list

Mitigate potential risks

Open up perspectives

2–4 hours for 3–5 risk zones

Interdisciplinary team

Senior stakeholders

© 2021 Thoughtworks

Our perspective

- Build the checklist into your product design requirements
- Distribute your top questions to your leadership and advisors to get their input
- Post the questions in forums or places where they will be kept top of mind
- Collect resources and seek out experts around your top risk areas to inform your strategy and design
- Revisit the checklist whenever you're starting development or planning to scale a new product or service

Explore further:

www.ethicalOS.org

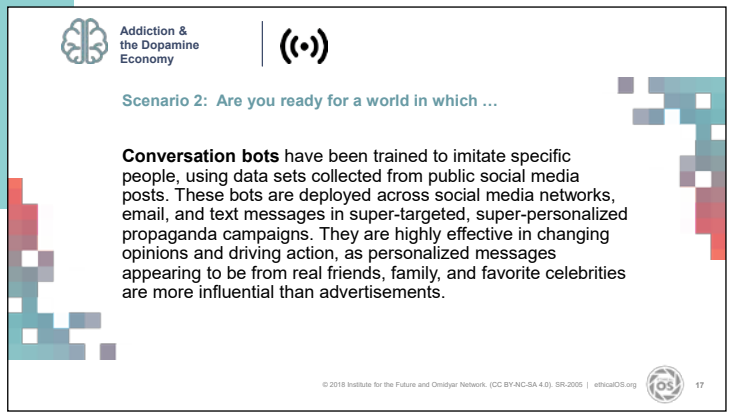
EthicalOS is an initiative by the **Omidyar Network's Tech and Society Solutions Lab**

https://omidyar.com/responsible-technology-2/

Ethical OS (continued)

Remember: Don't get caught up in whether a scenario is likely or even possible. Just pick one and go with it.

> When something of massive consequence happens that no one predicted, we often say it was simply unimaginable. But the truth is, nothing is impossible to imagine. When we say something was unimaginable, usually it means we failed to point our imagination in the right direction."
>
> *—Jane McGonigal, Institute for the Future*



How does it work?

1. Begin by choosing a technology, product or feature you're working on.
2. Identify the risk zones and questions related to each zone that is most relevant to the team and the technology you've chosen. [Use this checklist to get started with questions to ask the team.]
3. Use the questions to kickstart conversations within the team or organization.
4. Ask: "How can we start to correct or mitigate these risks?"
5. Document and begin to write out actions that could then be prioritized. Also consider at this stage:
   - What are the hardest or biggest threats?
   - Strategies that could help mitigate
   - Who needs to come on board before moving forward?

< Return to methods list

© 2021 Thoughtworks

Failure Modes & Effects Analysis

By American Society for Quality (ASQ)

What is it?

FMEA is a risk-based approach used by designers and engineers to design, build and run more resilient, reliable systems. It uses qualitative descriptions and quantitative scoring to calculate failure criticality and risk priority of potential failures (security, privacy, availability, performance, ethics) for analysis and prevention. It helps identify and prevent failures in multiple domains.

Why use it?

Human beings are not very good at facing possible catastrophes and not very creative in considering how the things we design might fail. The FMEA process helps by defining all the possible 'bad' outcomes and potential failures and helps us to determine the way in which each failure might happen so that we can prevent it.

When? When product or service concepts are being developed and whenever changes or decisions needs to be made. Initial run is four to eight hours, updates are about two hours. Moderate to difficult (requires a facilitator with experience)

With whom? Entire team (tech, design, BA, QA, SMEs); three to six is ideal, not more than six.

Failure	S	P	D	RPN
System reminds a user of a now-deceased relative	3	2	7	**42**
System recommends a car advertisement to someone who does not drive	2	5	3	**30**
System raises interest rates on user based on automated credit scoring algorithm	5	3	1	**15**

< Return to methods list

Mitigate potential risks

Initial run for 4–8 hours

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

It's a general method coming from the engineering sector and follows general principles from many other tools (e.g. agile threat modeling). However, while the core principles are a strength, it it needs to be further adjusted to work in your own process and context. Tools and templates exist but the language needs updating to work in the software environment.

Explore further:

https://asq.org/quality-resources/fmea **provides downloadable resources, detailed guides and examples.**

Originally produced and used by the U.S. Armed Forces and then the U.S. National Aeronautics and Space Administration (NASA) - the method is now supported by the **American Society for Quality**.

Failure Modes & Effects Analysis (continued)

Failure	S	P	D	RPN
System reminds a user of a now-deceased relative	3	2	7	**42**
System recommends a car advertisement to someone who does not drive	2	5	3	**30**
System raises interest rates on user based on automated credit scoring algorithm	5	3	1	**15**

How does it work?

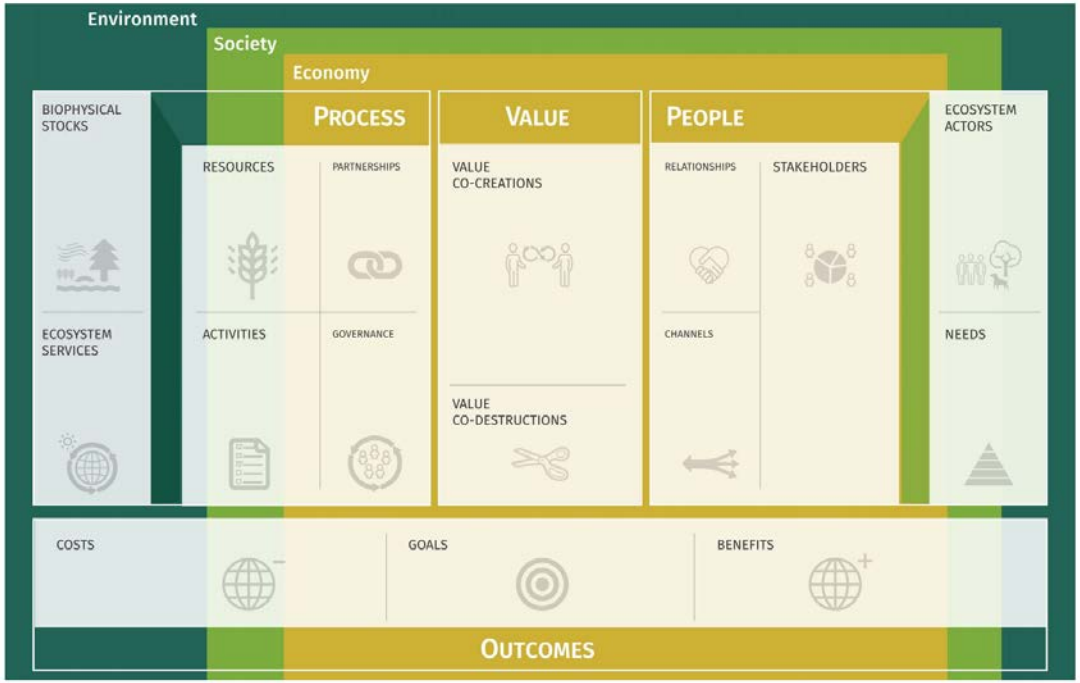
1. Define potential failure modes — describe how the system can fail.
2. Understand consequences of failures — the impact that the failures have on its users.
3. Identify possible causes and ratings — the possible mechanisms by which the failures occur, and the probability and detectability of the failures or causes.
4. Evaluate risk criticality and priority — identify appropriate designs and controls to mitigate or prevent the failures.
    - Can be done with Mural-type canvas and a Google Docs worksheet
    - Part of understanding a system's overall risk profile
    - It's important to use prompts and examples to facilitate thinking at each step along the way

< Return to methods list

© 2021 Thoughtworks

Flourishing Business Model Canvas

By Flourishing Enterprise Institute



What is it?

The Flourishing Business Model Canvas is an extension of the Business Model Canvas. It broadens conversations beyond feasibility, viability and desirability of a business, product or service to also consider economic, social and environmental aspects.

The canvas promotes collaboration with a diverse range of stakeholders and values through 16 questions – enabling alignment on key strategic decisions and create a common language.

Why use it?

- To understand the broader economic, social and environmental impacts when evaluating new business, product or service opportunities
- If you are comfortable working with Business Model Canvases this canvas is a little more complex due to the increased scope

When? In the early stages of your project or product lifecycle

With whom? The number of people who will need to participate will depend highly on the complexity and potential stakeholders who may be affected. Both internal and external stakeholders should be consulted, for example customers, community and environmental advocacy groups

< Return to methods list

Unpack stakeholder values

Open up perspectives

1–2 hours initially

Senior stakeholders

© 2021 Thoughtworks

Our perspective

The Flourishing Business Model Canvas provides a useful tool to consider and take into account potential economic, social, and environmental impacts of a business holistically. Identifying and engaging all affected stakeholders to validate thinking might be the most challenging aspect of this method.

Explore further:

http://www.flourishingbusiness.org/the-toolkit-flourishing-business-canvas/

Flourishing Business Model Canvas (continued)

Simon Robinson and Maria Moraes Robinson from Holonomics developing new flourishing business models with executives from the energy sector in Brazil.

How does it work?

The tool can be used in a similar way as the Business Model Canvas, and helps to not only focus on 'doing well', but also on 'doing good'.

- The canvas could initially be completed in a two to three hour collaborative workshop. It can then be continuously evolved through lean experiment and ongoing stakeholder validation
- It serves as a key tool to rapidly prototype and iterate on their business models
- The canvas allows businesses to start with their core values, striving to be flourishing entrepreneurs, and then allows them to design and test potential models

A case study of Dear Green Coffee putting the Flourishing Business Model into practice can be found in this blog post.

The Flourishing Business Canvas was created by the team members of the Flourishing Enterprise Innovation Toolkit Project © Antony Upward / Edward James Consulting Ltd., 2014 All Rights Reserved www.FlourishingBusiness.org Images used with permission. If you use the canvas, get in touch with them about licensing. Workshop photos © Holonomics, www.holonomics.co.

< Return to methods list

InterpretML

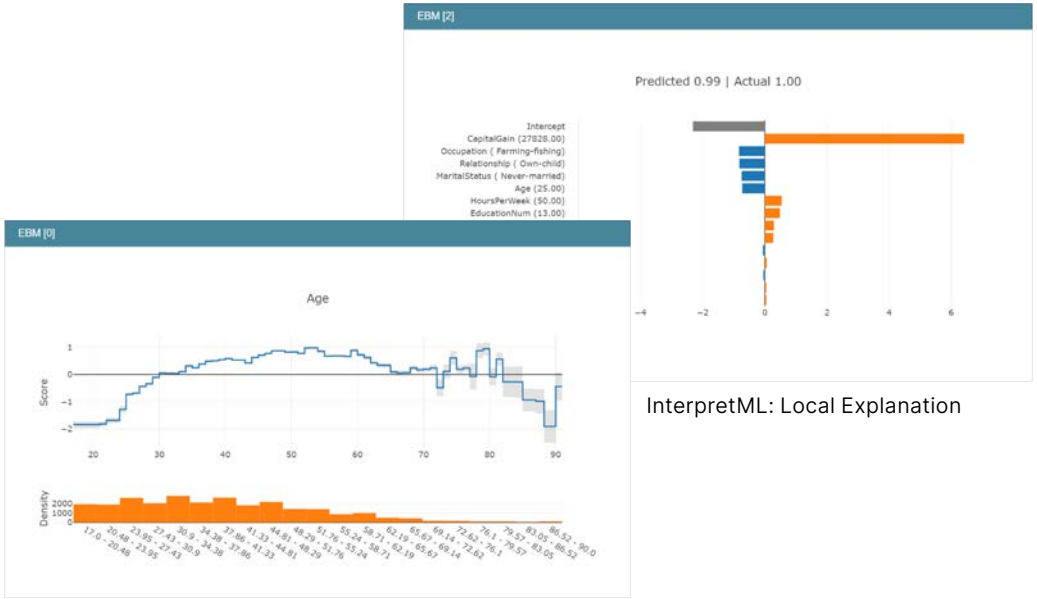
By Microsoft Research

What is it?

InterpretML is an open-source library that enables stakeholders to understand and explain the predictions of machine learning models. Its goal is to debunk the trade-off between predictive power and explainability. InterpretML provides state of the art methods (extending beyond LIME and SHAP) for demystifying an existing model's behavior, in addition to building brand new, transparent models.

InterpretML capabilities:

- Build 'glassbox' models that are explainable from the ground up and even editable by domain experts
- Users can also bring their own blackbox models and conduct post hoc explainability analysis
- Develop explanations for overall (global) model behavior as well as individual (local) predictions
- Drill down into subsets of predictions and conduct counterfactual (what-if) experiments
- Incorporate complementary frameworks for AI fairness such as Fairlearn and AIF360



InterpretML: Local Explanation

InterpretML: Explainable Boosting Machine (EBM) analysis

Why use it?

For developers and data scientists, ML interpretability is often crucial for debugging models, ensuring responsible predictions, and even protecting against adversarial attacks. For management and regulators, these tools help clarify product trustworthiness and support auditing for regulatory compliance.

When? Especially vital for high-risk applications (e.g. safety-critical or impacting social inequality).

With whom? Team members with Python knowledge and context about the data.

< Return to methods list

Mitigate potential risks

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

InterpretML makes Responsible AI convenient and accessible to stakeholders without advanced statistics knowledge, while still remaining technically rigorous.

We would particularly like to see modern glass-box models coming to the forefront, with the hope that AI performance and explainability no longer have to be necessarily antithetical.

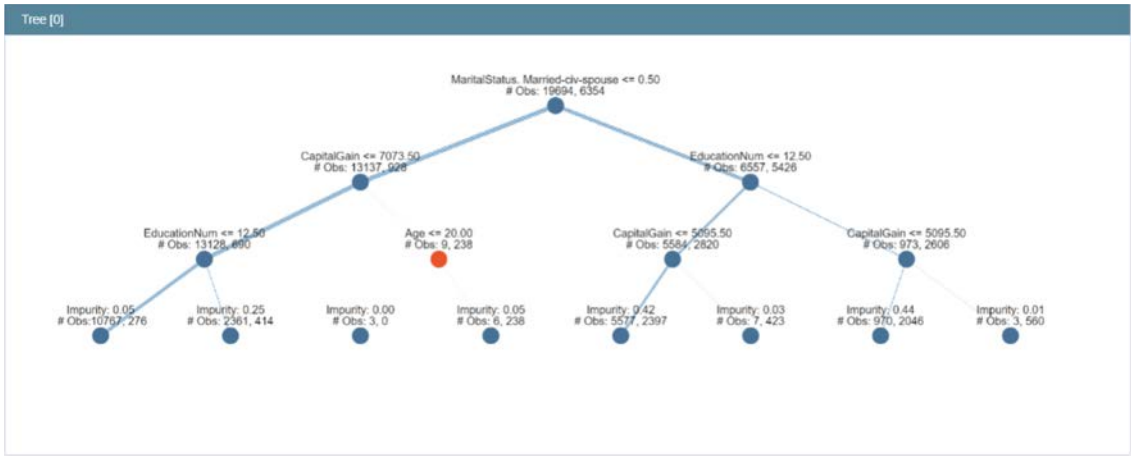
As interpretability can be relevant to issues of fairness, we recommend adding depth to your analyses and even algorithmically mitigating unfairness by integrating with fairness frameworks such as Fairlearn where applicable.

Complementary Tools: Fairlearn, IBM AI Fairness 360, PRESC, Google What-If Tool

Explore further:

https://interpret.ml

Interpret ML (continued)



InterpretML: Investigating the impact of a feature (Capital Gains in this example) for predicting the income of an individual. Data sample proportions are also automatically displayed for each rule.

How does it work?

1. Follow the Getting Started guide for InterpretML.
2. Consider *starting with* InterpretML's glass-box models such as an Explainable Boosting Machine (EBM). Some benchmarks have shown that EBMs are even comparable to popular high-performance models such as XGBoost!
3. Alternatively, bring your own arbitrary model then run post-hoc black-box explainers (e.g. SHAP Kernel Explainer) to analyze model behavior.
4. Analyze local and global explanations.
5. Using the InterpretML dashboard, conduct what-if investigations interactively by editing data points and visualizing new outcomes and explanations.

InterpretML is licensed by **Microsoft** under the MIT License

Analysis and images based on the Adult Income Dataset from Dua, D. and Graff, C. (2019). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

< Return to methods list

Materiality Matrix Assessment

What is it?

Materiality is the principle of identifying the social, economic and environmental issues that matter most to your company and your stakeholders. A materiality matrix will help with the visualization of the findings and insights of a materiality assessment.

A common way of organizing the matrix includes plotting the economic, social and environmental topics considering their importance to external stakeholders and their impact on the business (internal stakeholders).

Why use it?

The exercise is aimed at engaging stakeholders to find out how important specific economic, environmental and social issues are to them. The insights gained can then be used to guide strategy and communication, and tell a more meaningful sustainability story.

When? It can be particularly valuable at the start of a new project or product, or can be used as a recurring assessment, e.g. quarterly or yearly.

With whom? Stakeholders and the internal cross functional team.

It helps provide a holistic overview of the impact the business has not only in an economic context, but also on social and environmental issues.

UPS Materiality Matrix

< Return to methods list

Unpack stakeholder values

Up to 4 hours initially

Interdisciplinary team

Senior stakeholders

© 2021 Thoughtworks

Our perspective

The tool is most valuable when the company can get insights from both internal and external stakeholders. Create a list of relevant stakeholder groups who can provide a meaningful perspective on your company's sustainability strategy. Take into account both internal (i.e., executive leadership, regional managers, employees) and external contacts (i.e., trade associations, key customers, NGOs) so you can evaluate a wide range of perspectives across the value chain.

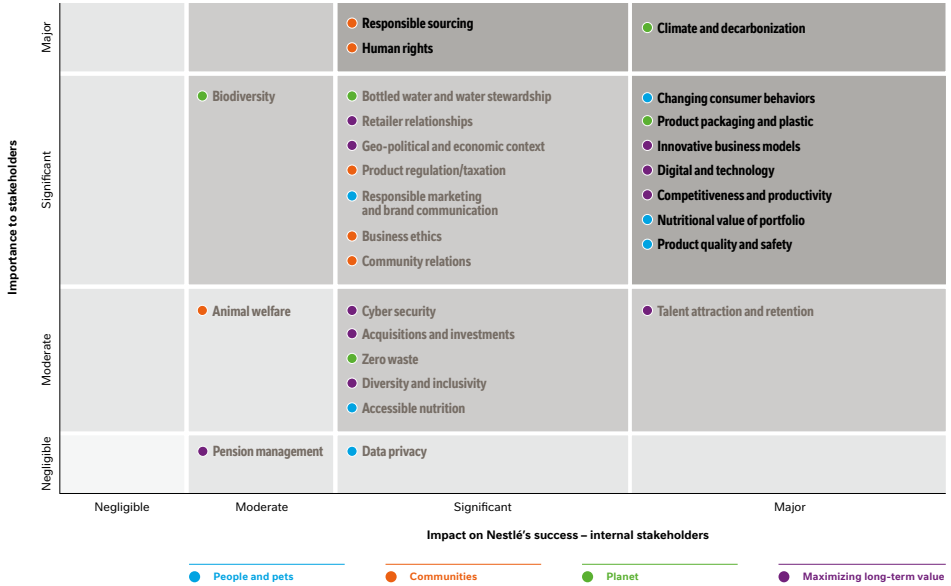
Explore further:

Detailed guides available

https://us.anteagroup.com/en-us/blog/7-basic-steps-conducting-successful-materiality-assessment

https://www.stern.nyu.edu/sites/default/files/assets/documents/NYUSternCSBSustainabilityMateriality_2019_0.pdf

Materiality Matrix Assessment (continued)



Nestlé Materiality Matrix

How does it work?

1. Start identifying your internal and external stakeholders who can provide a meaningful perspective on the company's sustainability strategy.
2. Identify and prioritize which sustainable indicators you want to use. For example: Economic (e.g., profit, revenue, company turnover); Social (e.g., labor statistics, community impact) and Environmental (e.g., water consumption, waste management).
3. Design your Materiality Survey. The stakeholders will rate the importance and impact of each sustainable indicator on a numerical scale (e.g.; 1–5 or 1–10).
4. Launch your survey and start collecting insights.
5. Create the Materiality Matrix based on the survey responses.
6. Share your results and put the insights into actions.

< Return to methods list

© 2021 Thoughtworks

Responsible Strategy

By Doteveryone (now Open Data Institute)

What is it?

Seven questions that can be used as a starting point for decision-making and objective setting that will move your organizational, team or product behavior closer to the values you want to uphold. Facilitates how values might become real, by translating them into principles and measurable KPIs and cross-checking for unintended consequences.

A useful questionnaire that can carve out fundamental building blocks for other discovery and delivery methodologies like Lean Value Tree or OKR.

Why use it?

Understanding, clearly articulating, and committing to your organizational values is the first step to better business planning and more responsible metrics. Principles and KPIs can provide a guide for the collective responsibility of those within the organization for the products they put into the world.

When? Project start (e.g. add-on to normal vision and mission session)

With whom? Team and management

Use as a starting point for decision-making and objective-setting to move closer to the values you want to uphold in the organization and/or project.

< Return to methods list

Unpack stakeholder values

1–2 hours

Interdisciplinary team

Senior stakeholders

© 2021 Thoughtworks

Our perspective

- Creates deep understanding of underlying organizational principles and values
- Groundbreaker for further facilitation
- Identify KPIs connecting progress on vision to value
- Good to include early in the process as well as with key-stakeholders
- Be aware, the open question format needs to be structured and condensed

Explore further:

Question sheet: https://www.tech-transformed.com/download/1008/

Review your values and metrics https://www.tech-transformed.com/strategy/

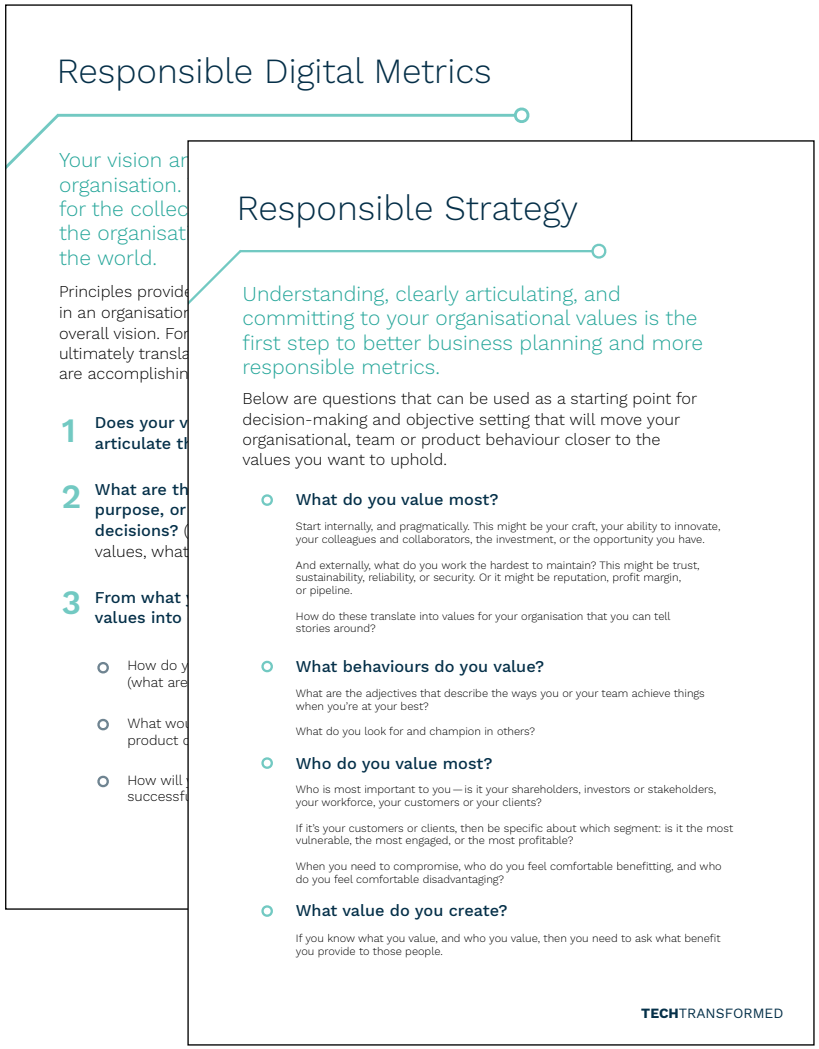
Produced by **Doteveryone** (now maintained by the **Open Data Institute**).

Attribution-Non-Commercial-Share-Alike 4.0 International (CC BY-NC-SA 4.0);

Responsible Strategy (continued)

How does it work?

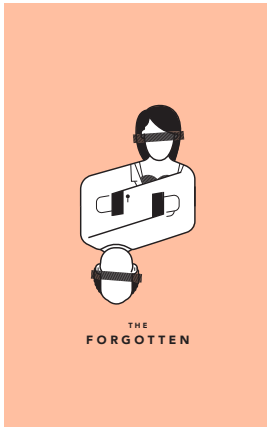
1. Most likely integrate into inception phase
2. Either do one-to-one interviews or give as “homework” to team
3. Collect answers in group session and consolidate
4. Document and make transparent going further in the project to underpin decisions
5. Review midway through the project.

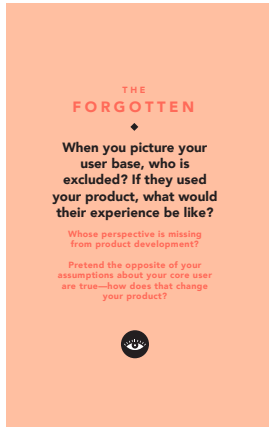


< Return to methods list

Tarot Cards of Tech

By Artefact





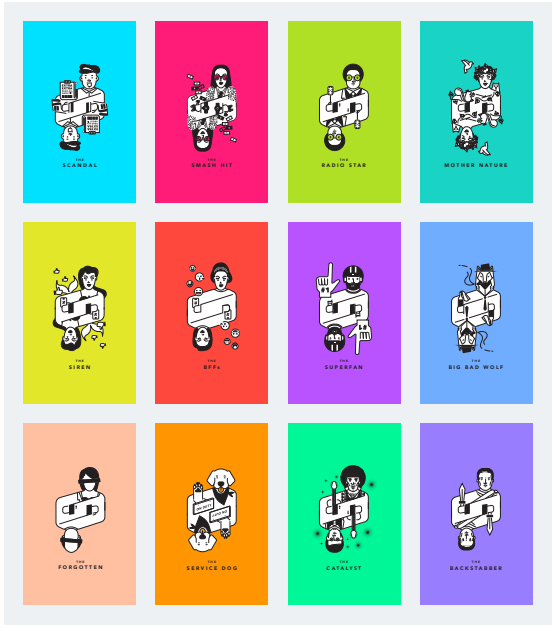
What is it?

The Tarot Cards of Tech are a brainstorming exercise to encourage creators to think about the true outcomes technology and products can create, from unintended consequences to opportunities for positive change. Each card contains a provocation, covering topics such as:

Usage. How could your product impact cultural habits and relationships? How could your product be used in ways (or by people) you haven't considered?

Equity and access. What users are excluded? What would their experience be like?

Scale and disruption. What would using your product 'too much' look like? What would you change if your client was Mother Nature?



Why use it?

To expand your thinking on the impact your product / technology can have from differing perspectives, to drive teams to make intentional decisions.

When? This tool would be great at early stages of product ideation (e.g. during an inception). Could also be used at the 'inception' and 'epic' level.

With whom? Include the whole team — including stakeholders.

< Return to methods list

Open up perspectives

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

It's a short session, easily possible, and the cards provide a good range of provocations. At times the cards could do with more clarity but do well to trigger conversations. Ensure there's a balance between 'worst-case scenarios' and 'good opportunities' — even though the 'evil' cards like Backstabber seem to generate more thinking and conversation.

Explore further:

(cards available)
http://tarotcardsoftech.artefactgroup.com/

Produced by **Artefact**.

Tarot Cards of Tech (continued)

How does it work?

This is a brainstorming exercise, intended to generate conversation so teams think bigger picture and longer term about the products they put out into the world.

1. Download and print the cards.
2. Break into smaller groups - we divided up the cards among the groups by theme.
3. Pick a card and read to the group - card contains a provocation to help teams see a product from different perspectives.
4. Discuss the questions it poses, taking notes (for playback!)
5. Share back with broader group, and remember to capture any actions out of the discussion.

You may want to consider:

- Breaking into small groups and holding a playback discussion
- How will you take actions based on your findings — make a plan
- Pick the most relevant cards and do a deep dive

How could you apply this?

- During design process, or persona creation
- For strategy or new ideas, especially during design sprints
- Brainstorming negative consequences
- Out of the box thinking / out of the ordinary
- Noticing blind spots
- Identifying edge cases/behavior

Workshop at the Humane Tech Berlin Group

< Return to methods list

© 2021 Thoughtworks

Unintended Consequences

By Hyper Island

What is it?

In this workshop, groups examine the unintended consequences of new technologies and use those to inspire potential business opportunities. The workshop builds on the idea of “unintended consequences” as a starting point for generating ideas. It looks beyond the common understanding of new technologies, challenging participants to discover unexpected potential and how it might be harnessed. For instance, Instagram was created as a virtual photo album for smartphones, but through the human connection between the people who use it, it has the potential to affect something as deep as our grieving and healing process, as a recent example of this widowed writer shows.

Why use it?

Use this tool to demonstrate what happens when products are not built responsibly. This tool will convince your team to consider ethics, create sense of urgency to look at potential areas in your solutions and make ethics a priority in your organization.

< Return to methods list

Open up perspectives

1–2 hours per activity

Interdisciplinary team

© 2021 Thoughtworks

Our perspective

- You can run this workshop to uncover some of the possible ramifications of your new business plan or idea
- To do that, in step two, skip listing new technologies or businesses, and explain your solution to the group, including your identified challenges
- After this workshop, many patterns emerge regarding how technology or business can potentially harm society. Participants generally reflect on their new business model, but not on unintended consequences of them. As there are not many tools that bring this discussion to the fore, particularly in connection with business models, this is a big plus

Explore further:

http://toolbox.hyperisland.com/unintended-consequences

Produced by **Hyper Island**.

Unintended Consequences (continued)

How does it work?

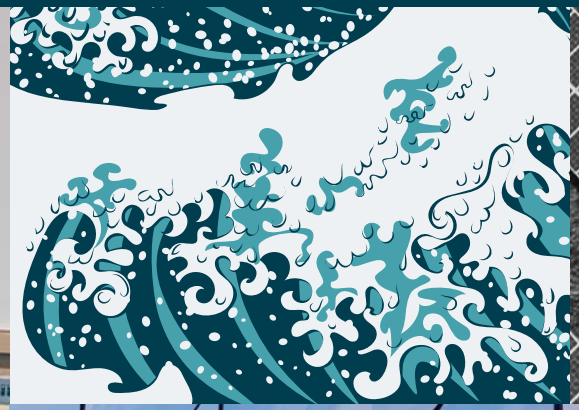
1. Begin by organizing the team into small groups of three to five. Introduce the purpose of the activity by sharing an example or getting the groups to share their own, relating stories about unintended consequences in technology.
2. Give groups five minutes to create a list at least five new technologies or businesses, such as Netflix, Instagram, Google Maps, etc. To add a level of challenge, have participants list only technologies widely used by demographics other than their own. Write them down on post-its (one per post-it).
3. Next, ask participants to brainstorm: for each of the new technologies or businesses they should speculate on at least three unintended consequences. For example Netflix was just supposed to be an easy and convenient streaming service, but it has led to new behaviors like 'binge watching'. Encourage the participants to think creatively and stress that there are no right answers. Ask them to write them down on post-its (one per post-it). After 10 minutes ask each group to share their ideas.
4. In smaller groups, for each of the unintended consequences, suggest a new possible business model. For example Netflix could invent a new business model where users can watch shows for free if they watch only one episode each day, or pay a premium for their binge behavior. When ideating, welcome all ideas as there are no bad ideas. After 12 minutes ask the group to gather and share their most interesting new business models with the group.
5. Ask the participants to reflect on the most unexpected and most viable new business model and why? They can also reflect what this exercise taught them about ideation and new business models.

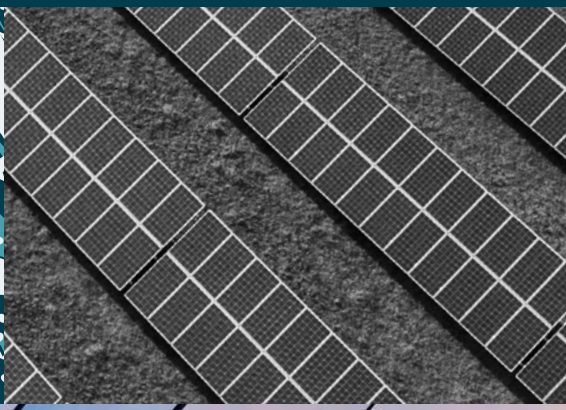
< Return to methods list

© 2021 Thoughtworks

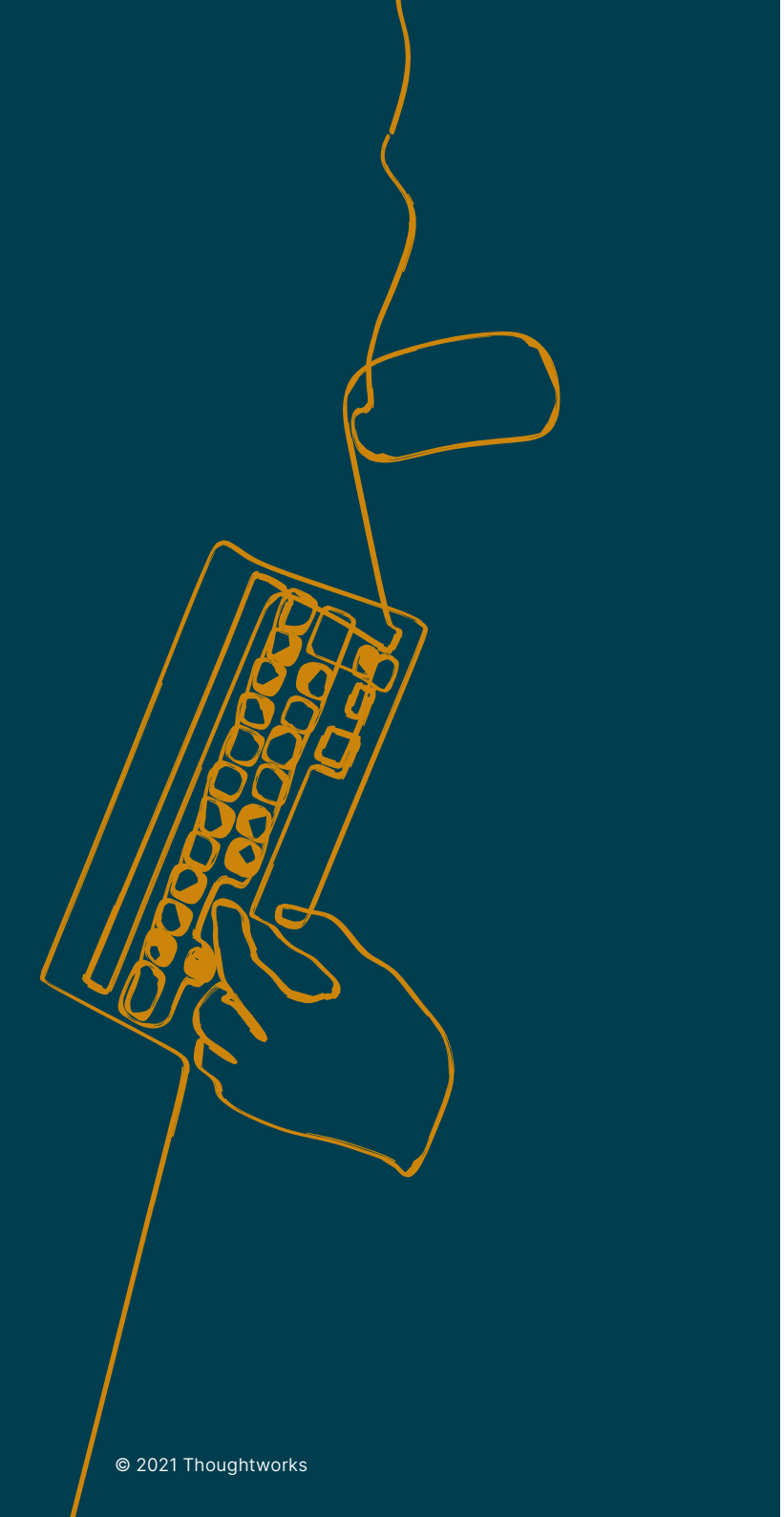
Continuing the Responsible tech journey

There are many more approaches and methods than we've covered in-depth. Here, we showcase some of them.





© 2021 Thoughtworks



Promising approaches

Polarity Thinking

Most ethical practice will lead to tensions between competing values. These polarities (sometimes also called wicked problems, complex problems, dilemmas, paradoxes, tensions etc.) cannot be solved like a conventional “problem,” i.e. working out which of the two options is the “solution” and then going for it. Instead, they need to be managed in a “different” way. Methods and mindsets based on Polarity Management, which was made popular by Barry Johnson in the nineties, can help with this.

Explore further:

Polarity Thinking* Polarity Mapping Overview by Neesa Sweet
http://braidedrivergroup.com/wp-content/uploads/2011/04/Polarity-Thinking-Overview.pdf

Barry Johnson (2014). *Polarity Management: Identifying and Managing Unsolvable Problems*. HRD Press.

Hands on POLARITY THINKING - Leveraging Difference for Innovation by Coreative
https://www.wearecocreative.com/pot-of-gold

© 2021 Thoughtworks

Promising approaches

Humane Design Guide

The Human Design Guide by Center for Humane Technology explores how we can build technology that doesn't exploit but does supports us. A worksheet to help you take meaningful steps towards designing a more humane product and to identify where investing in a deeper understanding of human nature will yield further benefits. It leads you through an assessment of the six human sensitivities and helps you take action with your team. It has two pages, one to assess the current state of your product and one to ideate on.

While the worksheet isn't the answer to all problems, the whole field around Humane Design is an important field that makes one more aware on how we want to define our relationship with technology.

Explore further:

https://humanetech.com/designguide/

Humane by Design (https://humanebydesign.com/) provides a resource that provides guidance for designing ethically humane digital products through patterns focused on user well-being.

© 2021 Thoughtworks

Promising approaches

Speculative design / provocations

This is especially interesting for designers. Normally, design and designers are problem solvers, focused on making a product better or more beautiful, or to make a process more efficient, but here, instead of solving problems, they pose them.

Design is used to imagine how that future might be entirely different, to make scenarios experienceable that help to illuminate (or make visible) moral, ethical, political and aesthetic problems. How does this idea work for a black guy? A white woman? For a child in a wheelchair? For a very rich — or for low-income family? Who else is going to be touched by this innovation?

Explore further:

https://uxplanet.org/can-speculative-design-make-ux-better-design-trend-4-4-ce8d13148e5d

© 2021 Thoughtworks

Promising approaches

Bracketing

Bracketing is a method used in qualitative research to mitigate the potentially deleterious effects of his or her presuppositions, biases, assumptions, theories, or previous experiences.

This method can also be used during the development process. Designers and developers make notes, like architecture decision records (ADR), whenever they sense that one of their own biases might have influenced their decision. Such a list can then later be be used for a wider discussion as well as to check and alter algorithmic outcomes.

Explore further:

https://www.researchgate.net/publication/257924681_Bracketing_in_Qualitative_Research

© 2021 Thoughtworks

Promising approaches

Participatory design methodology

Previous examples have mostly been tools, whereas participatory design is more of a methodology and philosophy of how to design. The main takeaway is participants being far more involved — more than just asking them what they want, in PD users are owners of the content and process. This recognizes that users are the experts, we are facilitators helping them get their answers and ultimately they should have more power in product design.

Explore further:

https://en.wikipedia.org/wiki/Participatory_design

© 2021 Thoughtworks

Other emerging approaches to explore

Our guide is not a definitive list, and there are a growing number of methods and tools available. **Continue your journey of responsible and ethical tech development by exploring these additional tools and resources.**

Click on the links to explore

- AI Blindspot Cards by MIT Media Lab — identifying blindspots during planning, building and deploying of an AI product
- AI Ethics Cards by IDEO — a collection of questions, methods and examples that help interdisciplinary teams stay human-centered
- Bracketing in research — used in qualitative research to mitigate the potentially deleterious effects that may taint the research process
- Ethics Canvas by the ADAPT Centre for Digital Content Technology — adapted from Alex Osterwalder's Business Model Canvas
- Equity-Centered Community Design Field Guide by Creative Reaction Lab — an equity-centered problem solving process focused on co-creating with the community
- Futures Wheel by Jerome C. Glenn — brainstorm the tertiary unintended or unforeseen effects of a product
- Inverted Behaviour Model by Katherine M. Zhou — invert BJ Fogg's Behaviour Model by designing for the unintended behavior
- Ethical and Rights-Based AI Approaches by BKC for Internet & Society — a map of the various AI principles and guidelines
- Needs & Collateral by Katherine M. Zhou — helps to identify negative collateral and undesirable side-effects for users
- Privacy by Design by Ann Cavoukian — it calls for privacy to be taken into account throughout the whole engineering/design process
- Green Patterns: Helping Users Make More Sustainable Choices by Tim Frick
- Systemic Design Framework by UK Design Council

© 2021 Thoughtworks

Further reading

The following online resources offer more valuable perspectives on the field of responsible and ethical technology, and plenty of food for thought.

Click on the links to explore

- Ethical and Rights-Based AI Approaches by BKC for Internet & Society — an overview map of Ethical AI approaches
- IEEE's Ethically Aligned Design: Prioritizing Human Wellbeing with Autonomous and Intelligent Systems
- Guide to Responsible tech: How to Get Involved & Build a Better Tech Future by All Tech is Human
- Teaching Responsible Computing Playbook by Mozilla
- The Business Case for AI Ethics by All Tech is Human
- A general Ethical & Humane-Tech Reading List

Acknowledgments

This book wouldn't have been possible without the work, wealth of knowledge, and expertise of many individuals, organizations and partners – thank you.

Further, we would like to thank everyone at Thoughtworks who discussed, trialled, added and reviewed Responsible Tech Tools and the documentation around them. This is truly a collection of many people's experiences, thank you all!

© 2021 Thoughtworks

Continue the conversation

If you've tried any of these tools we would love to hear how it went. Send us an email, or use the form to suggest another method that you think we should include here. If you need some help, or have questions, we would be happy to set up a workshop to help you in a more tailored way, our goal is to further support and catalyze successes in this field.

Contact us by email:
RTteam@Thoughtworks.com
Use this **form to submit another method or tool**.

/thoughtworks

© 2021 Thoughtworks