ETHICS IN TECH PRACTICE:

A Toolkit

MARKKULA CENTER FOR APPLIED ETHICS
at Santa Clara University

AN ETHICAL TOOLKIT FOR ENGINEERING/DESIGN PRACTICE

AUTHOR:
SHANNON VALLOR, REGIS AND DIANNE MCKENNA PROFESSOR, SANTA CLARA UNIVERSITY

The tools below represent concrete ways of implementing ethical reflection, deliberation, and judgment into tech industry engineering and design workflows.

Used correctly, they will help to develop ethical engineering/design practices that are:

- ***Well integrated*** into the professional tech setting, and seen as a natural part of the job of good engineering and design (not external to it or superfluous)
- ***Made explicit*** so that ethical practice is not an 'unspoken' norm that can be overlooked or forgotten
- ***Regularized*** so that with repetition and habit, engineers/designers/technologists can gradually strengthen their skills of ethical analysis and judgment
- ***Operationalized*** so that engineers/designers are given clear guidance on what ethical practice looks like in their work setting, rather than being forced to fall back on their own personal and divergent interpretations of ethics

Each tool performs a different ethical function, and can be further customized for specific applications. Team/project leaders should reflect carefully on how each tool can best be used in their team or project settings. Ask questions like the following:

- What part of our existing workflows would this tool naturally fit into? If none, where in our workflows could we make a good place for it?
- What results do we want this tool to help us achieve? What risks do we want its use to mitigate or diminish?
- How often should this tool be used, in order to achieve those goals?
- Who should be involved in using this tool, and who on the team should be assigned responsibility for overseeing its use?
- In what ways should use of the tool, and the outcomes, be documented/evaluated?
- How will we reward/incentivize good use of these tools (for example, in performance reviews) so that employees are strongly motivated to use them and do not seek to avoid/minimize their use?
- What training, if any, do employees need in order to use these tools properly, and how will we deliver that training?

Each of the seven tools is summarized on the next page, with fuller descriptions of each and examples of possible implementations on the pages that follow.

TOOL 1: ETHICAL RISK SWEEPING: *Ethical risks* are choices that may cause significant harm to persons or other entities with a moral status, *or* are likely to spark acute moral controversy for other reasons. Failing to anticipate and respond to such risks can constitute *ethical negligence*. Just as scheduled penetration testing and risk sweeping are standard tools of good cybersecurity practice, ethical risk sweeping is an essential tool for good design and engineering practice.

TOOL 2: ETHICAL PRE-MORTEMS AND POST-MORTEMS: While Tool 1 focuses on individual risks, Tool 2 focuses on avoiding ***systemic*** ethical failures of a project. Many ethical disasters in design and engineering have resulted from the *cascade effect*: multiple team failures that in isolation would have been minor, but in concert produced aggregate ethical disaster. Thus we need a tool geared toward the dynamics of *systemic* design failure, something that ethical pre- and post-mortems are suited to offer.

TOOL 3: EXPANDING THE ETHICAL CIRCLE: In most cases where a technology company has caused significant moral harm due to ethical negligence, the scope of the harm was not anticipated or well-understood due, at least in part, to forms of cognitive error that lead designers and engineers to ignore or exclude key stakeholder interests. To mitigate these common errors, design teams need a tool that requires them to 'expand the ethical circle' and invite stakeholder input and perspectives beyond their own.

TOOL 4: CASE-BASED ANALYSIS: Case-based analysis is an essential tool for enabling ethical knowledge and skill transfer across ethical situations. It allows us to identify prior cases that mirror our own in key ethical respects; to analyze the relevant parallels and differences; to study adopted solutions and strategies, and their outcomes; and to draw reasoned inferences about which of these might helpfully illuminate or carry over to our present situation.

TOOL 5: REMEMBERING THE ETHICAL BENEFITS OF CREATIVE WORK: Ethical design and engineering isn't just about identifying risks and avoiding disaster; it's about a *positive* outcome: human flourishing, including that of future generations, and the promotion of healthy and sustainable life on this planet. Too often, other goals obscure this focus. To counter this, it helps to implement a workflow tool that makes the ethical benefits of our work explicit, and reinforces the sincere motivation to create them.

TOOL 6: THINK ABOUT THE TERRIBLE PEOPLE: Positive thinking about our work, as Tool 5 reminds us, is an important part of ethical design. But we must not envision our work being used only by the wisest and best people, in the wisest and best ways. In reality, technology is power, and there will always be those who wish to abuse that power. This tool helps design teams to manage the risks associated with technology abuse.

TOOL 7: CLOSING THE LOOP: ETHICAL FEEDBACK AND ITERATION: Ethical design and engineering is never a finished task—it is a loop that we must ensure gets closed, to enable ethical iteration and improvement. This tool helps to ensure that ethical initiatives and intentions can be *sustained* in practice, and do not degrade into 'ethical vaporware.'

TOOL 1: ETHICAL RISK SWEEPING

Ethical Risks are choices that may cause significant harm to persons, or other entities/systems carrying a morally significant status (ecosystems, democratic institutions, water supplies, animal or plant populations, etc.) *or* are likely to spark acute moral controversy for other reasons.

Ethics in technology design and engineering often begins with seeking to ***understand*** the moral risks that may be created or exacerbated by our own technical choices and activity; only then can we determine how to reduce, eliminate, or mitigate such risks.

In the history of design and engineering, many avoidable harms and disasters have resulted from failing to adequately identify and appreciate the foreseeable ethical risks. Such failures are a form of ***ethical negligence*** for which technologists can be held responsible by a range of stakeholders, including those directly harmed by the failure, the general public, regulators, lawmakers, policymakers, scholars, media, and investors.

Why do foreseeable ethical risks get missed?

Ethical risks are particularly hard to identify when:

- We do not share the moral perspective of other stakeholders
- We fail to anticipate the likely causal interactions that will lead to harm
- We consider only material/economic causes of harm
- We fail to draw the distinction between conventional and moral norms
- The ethical risks are subtle, complex, or significant only in aggregate
- We misclassify ethical risks as legal, economic, cultural, or PR risks
- We lack explicit, regularized practices of looking for them

How can we mitigate these challenges?

- Institute ***regularly scheduled*** ethical risk-sweeping exercises/practices to strengthen and sustain the team's ethical 'muscle' for detecting these kinds of risks
- Assume you ***missed*** some risks in initial project development phase; ***reward*** team members for spotting new ethical risks, especially ones that are subtle/complex
- Practice ***assessing*** ethical risk: which risks are trivial? Which are urgent? Which are too remote to consider? Which are remote but too serious to ignore?
- Treat just as you would cybersecurity penetration testing; 'no vulnerabilities found' is generally ***good*** news, but you don't consider it wasted effort. You ***keep doing it.***

Implementation Example: Company A makes wearable 'smart' devices for health and wellness. During company onboarding, employees complete a comprehensive half-day workshop highlighting the ethical norms and practices of the company; this includes introduction to the risk-sweeping protocol, among others. The onboarding workshop includes training on the *particular* risks (e.g., to safety, privacy, autonomy, dignity, emotional well-being, etc.) concentrated in the health and wellness sector, and the risks specific to wearable design in this sector (for example, GPS-enabled risks to locational privacy, the risk of obsessive self-monitoring in some populations).

At Company A, all project managers must implement risk-sweeping protocols at four stages of their workflow: 1) initial product proposal (the 'idea generation stage'), 2) the prototype stage, 3) the beta-testing stage, and 4) the post-ship quality assurance stage.

Each phase of risk sweeping involves a **mandatory** team meeting or its equivalent, in which each team member is expected to identify and present some risks; productive contributions to these meetings must be noted on performance reviews.

At one or more stages, project managers must seek *outside* input into the process to ensure that the risk-sweeping protocol is not constrained by groupthink or a 'bubble' mentality. For example, they may work with Marketing to ensure that input on possible ethical risks is sought from a diverse focus group, or they may seek such feedback from beta-testers, or from tech ethicists or critics willing to offer input/advice under an NDA.

Each phase of risk sweeping builds upon the last, under the assumption that one more significant risks may have been missed in a prior stage, or has newly emerged due to a design change or new use case. Ethical risks at each stage are identified, assessed, classified and documented, even if trivial or remote. Assuming the absence of any 'no-go' risks (those that would necessitate abandoning the project), risks that continue to be classified as significant must then be subjected to a *monitoring and mitigation* strategy.

TOOL 2: ETHICAL PRE-MORTEMS AND POST-MORTEMS

While the risk-sweeping protocol focuses on individual risks, this tool focuses on avoiding *systemic* ethical failures of a project. Many ethical disasters in engineering and design have resulted from the *cascade effect*: multiple team failures that in isolation would not have jeopardized the project, but in concert produced aggregate ethical disaster. Thus an ethical risk-sweeping protocol should be paired with a tool geared toward the dynamics of systemic design failure, something that ethical pre- and post-mortems are suited to offer.

The concept of a **post-mortem** is familiar; under certain circumstances, such as when a patient dies under medical care in a manner or at a time in which death was not expected, the medical team may be tasked with a review of the case to determine what went wrong, and if the death could have been reasonably anticipated and prevented.

By highlighting *missed opportunities, cascade effects, and recurrent patterns* of team failure, such exercises are used to improve the medical team's practice going forward. To encourage open sharing of information and constructive learning, documentation of team failures in post-mortems is, in many contexts, designed as a ***non-punitive*** process; the purpose is not to assign or apportion blame to, or punish individuals, as it would be in a judicial review, but to determine how the ***system*** or ***team*** failed to prevent such failures, and how improved procedures and protocols can enable better outcomes in the future.

A version of the very same process can aid in technical design and engineering settings. It can be enhanced with a ***pre-mortem*** protocol. Instead of *waiting* for ethical disasters to happen and then analyzing them, teams should get in the habit of exercising the skill of *moral imagination* to see how an ethical failure of the project *might* easily happen, and to understand the preventable causes so that they can be mitigated or avoided.

Team Post-Mortems Should ASK:
Why Was This Project an *Ethical* Failure?
What *Combination or Cascade* of Causes Led to the Ethical Failure?
What Can We *Learn* from This Ethical Failure that We Didn't Already Know?
What Team Dynamics or Protocols Could Have *Prevented* This Ethical Failure?
What Must We *Change* if We Are to Do Better Next Time?

Team Pre-Mortems Should ASK:
How Could This Project Fail for *Ethical Reasons*?
What Would be the Most Likely *Combined Causes* of Our Ethical Failure/Disaster?
What *Blind Spots* Would Lead Us Into It?
Why Would We *Fail to Act*?
Why/How Would We Choose the *Wrong* Action?
What Systems/Processes/Checks/Failsafes Can We Put in Place to *Reduce* Failure Risk?

Implementation Example: Company B makes massive multiplayer online video games. Five years ago, they had a very costly commercial failure of a game, 'Project Echo,' that injured their brand, wasted years of investment, and resulted in departures of some highly talented designers and other valued personnel. The failure had many *ethical* dimensions: the game was perceived by the gaming community and gaming media as a transparently exploitative 'pay to play' money-grab that, through its design choices, unfairly excluded or disadvantaged those players with less disposable income; it also unwittingly incentivized certain antisocial player behaviors that led to serious online and offline harms, and prevented the emergence of a healthy and growing player community. Finally, it included portrayals of certain social groups that were perceived by many, including vocal critics outside the gaming community, as insensitive and morally offensive.

Company B is determined to avoid this kind of disaster in the future.

They implement an extensive post-mortem of Project Echo, focusing on the systemic and cascading weaknesses of the design process that led to the outcome. They learn that each of the ethical risks were anticipated at several points in the design and development of the game, but due to poor communication between the creative, technical, and marketing teams, those worries were never addressed. They also learn that the game suffered from the company's lack of clear and consistent messaging to employees about its ethical principles and values; for example, how it regards 'pay to play' models, what kind of player communities it wants its games to foster, and how it wants its game narratives to fit within the broader ethical norms of society. Finally, they learn that team leaders had unwittingly set up perverse incentives that were meant to foster team 'cohesion,' but instead ended up rewarding careless design choices and suppressing the surfacing of worries or concerns about the risks created by those choices. The company seeks anonymous input from all ranks of employees on possible solutions, from which data they implement a number of changes to game design workflows and procedures to improve the ethical viability of future game projects.

They *also* implement a game design 'pre-mortem' requirement that must be executed jointly by the creative and production team leaders at the pre-production phase, in which team leaders incentivize their members to come up with multiple creative scenarios in which the project might fail. Technical and commercial failure risks are identified, but specifically *ethical* failure risks are explicitly required to be identified as well, and framed as such.

The pre-mortem process is supported by an addition to the company onboarding process, in which employees are presented with an overview of the company's ethical values, culture, and processes; provided with a review and discussion of the distinctive ethical risks and concerns that emerge in game design and development; given a conceptual framework and vocabulary for identifying such ethical concerns; and asked to review and discuss an ethical case study, such as the post-mortem of Project Echo.

TOOL 3: EXPANDING THE ETHICAL CIRCLE

In most cases where a technology company has caused significant moral harm, violated ethical norms in ways that damage internal morale and reputational standing, or invited aggressive regulatory oversight due to their ethical negligence, the scope of the harm was not anticipated or well-understood due, at least in part, to pernicious forms of:

- **Groupthink**: a social phenomenon in which the cognitive processes of a tight-knit group become too closely aligned, so that they begin to think 'in lockstep' and become unable to consider or accurately assess alternative perspectives other than those currently operating.

- **The 'Bubble' Mentality**: similar to groupthink, but caused not by a group's social dynamic, but by their demographic and cognitive similarities to one another; put in other terms, a cognitive and moral failure caused by a lack of sufficient diversity of life experiences, values, worldviews, identities, abilities, and/or personality styles. Environments in the tech industry, where teams may have very similar levels of educational attainment, many shared values and interests, common cultural assumptions and vocabularies, similar gender identities, ethnicities, age group, and physical abilities. Add to this the additional cohesion of a shared work culture and identity, and you have a breeding ground for a dangerous 'bubble' mentality. It takes a deliberate and concerted effort to counteract this phenomenon, in which 'good people' with 'good intentions' can easily make unethical decisions due to their insular cognitive view and its blindspots. This is why slogans like 'technology for social good' and 'making the world a better place' can be dangerous; they allow people operating within a bubble mentality to sincerely believe that they are acting ethically, when in fact they may lack cognitive access to the broader social realities they would need to understand in order to do so.

- **The 'Friedman Fallacy':** The economist Milton Friedman notoriously argued in the 1960's and 70's that companies, and employees acting on their behalf, are morally obligated *only* to maximize shareholder profit, and in no way responsible for considering the impact of their actions on the public interest—other than to stay within the 'rules of the game', i.e., the law. This view has been rightly criticized, not only for licensing grievous corporate harms to the public, but also for being anathema even to the moral foundations of capitalism outlined by Adam Smith and others, who tied the legitimacy of capitalism to the public good. Unfortunately, Friedman's fallacy is still taught in many business schools and other environments where all too often it is used to justify a company's deliberate or reckless disregard of the legitimate moral interests of affected stakeholders, or the public in general. The public, it must be noted, does not generally accept this fallacy. If a company knowingly poisons a local river with its toxic waste, but does so legally via a loophole in federal environmental regulations, the local residents do not shrug and say, 'well, of course, the company executives really had no choice - they *had* to give our kids cancer, after all, it would have been wrong to impose the costs of safe disposal on the shareholders!' Likewise, when a technology company leaves sensitive personal information unencrypted and wide open to hackers to save on the cost of security, or quietly sells it to third-parties with no restrictions on its use, and innocent people lose their life savings or safety as a result, no one accepts the Friedman Fallacy as an excuse, even if the law did not prohibit such actions.

These are grave and widespread causes of ethical failure in technology practice, but they can be addressed by explicit and deliberate measures designed to ***expand the ethical circle.*** That is, to ensure that the legitimate moral interests of the full range of stakeholders (people directly or indirectly affected by our actions) have been taken into account.

Expanding the ethical circle can be implemented in an explicit and regularized design exercise; also, it can and should be implemented at higher levels of corporate leadership with respect to the broader impact of the company's activities on society.

The exercise should pose questions such as the following, and invite explicit reflection upon the answers, as well as any active steps that should be taken as a result:

- Whose interests, desires, skills, experiences and values have we simply *assumed*, rather than actually *consulted*? *Why* have we done this, and with what justification?

- Who are all the stakeholders who will be directly affected by our product? How have their interests been protected? How do we know what their interests *really* are—have we *asked*?

- Who/which groups and individuals will be *indirectly* affected in significant ways? How have their interests been protected? How do we know what their interests *really* are—have we *asked*?

- Who might use this product that we didn't *expect* to use it, or for purposes we didn't initially *intend*? How does this expand/change the stakeholder picture?

- Who is at *substantial* risk of harm from our product, and how? How have we justified and mitigated this risk, and what have we done to procure the *informed and meaningful* consent of those at risk?
- Who are the people who will be least likely to purchase or use this product, but might have strong opinions about it anyway? Can those opinions be heard/evaluated by us?

Implementation Example: Company C is designing an app that aims to assist young people with autism in navigating social settings. The design team is highly motivated to help people with autism integrate into society more comfortably and effectively; it is clear to everyone involved that this is a morally noble aim, technically within reach, and worthy of pursuit on both ethical and commercial grounds.

However, Company C's internal ethical design protocols prompt them to expand the ethical circle in ways that will ensure that they do not fall victim to groupthink (being so collectively overtaken by their moral enthusiasm that no one thinks of any downsides or risks), a bubble mentality (after all, none of the designers are autistic, and all come from privileged economic backgrounds with similar educational and cultural experiences), or the Friedman Fallacy (having the noble aims of the project sidelined down the road by profit considerations that place the very users they were trying to help at risk).

During a team session early in the ideation process, one dedicated specifically to this tool, they begin to work through the questions above.

They determine quickly that they should consult with a range of young adults *with* autism before setting specific design goals. They realize that they have only *assumed* that the autism community would welcome or perceive a need for such an app, and that have not actually *asked* members of that community what their goals and interests might be in relation to their social experiences. They also identify a potential class of users that do not have autism but may seek out the app to help them manage general social anxiety, and must consider whether the design specifications should expand to features designed for those users. They also realize that trained professionals involved in the support of people with autism might have strong opinions about this app and its design, and that they would need to consider whether this app would be seen or used as a replacement for other modes of support.

After inviting a group of young autism activists and care professionals to a conversation, the design team is presented with a far richer set of ethical perspectives than they started with. They come to realize that several of their initial design goals and features might not be as helpful as they thought for all users, given the great diversity of forms of autism. They realize that they were operating with many stereotypes about autism, some more grounded in movies and television than in the real lived experiences of autistic people. They also realize that such experiences of being autistic can vary considerably according to other social factors, including cultural and economic differences. They also learn of some specific privacy and safety risks that many of the app's prospective users would want considered and addressed in the design.

They learn also that a significant subset of people with autism are resistant to interventions framed as 'treatments' or 'therapies'; they do not see autism as a 'problem' to be solved by technology, but simply a *difference* in a way of relating to the world. Many want a greater onus to be put on people in society *without* autism to be more socially receptive and accommodating to those with autism. Still, through the conversations the designers identify some areas of strong consensus about common social difficulties that some of their ideas for the app might help many users navigate; they decide to move forward with the app but in continued dialogue with the full range of stakeholders, who are invited to offer later input in the beta-testing and marketing stages to ensure that the app functions and is described in appropriate ways.

TOOL 4: CASE-BASED ANALYSIS

It is essential in ethical practice to be able to transfer ethical knowledge and skill across cases, so that we are not 'starting from zero' every time we analyze an ethical situation. The tool of case-based analysis is an essential and long-standing way of executing ethical knowledge and skill transfer.

The procedure of case-based analysis is fairly straightforward:

1. **Identify Similar or 'Paradigm' Cases that Mirror the Present Case**

   Where/when has a case relevantly like this one (in its ethical dimensions) occurred before? Which are 'clear' or 'paradigm' cases of the kind of ethical situation facing us?

2. **Identify Relevant Parallels Between/Differences Among All the Cases**

   In what ethically relevant respects is the present case *like* these paradigm cases? (For example, this case affects the same group of stakeholders, or introduces the same risks, or presents the same moral dilemma). In what ethically relevant respects is the present case *different* from the paradigm cases? (this time the stakes are lower, this time the law provides clear guidance, this time the public mood is different).

3. **Evaluate Choices Made and Outcomes of the Paradigm Cases**

   What was *done* in the paradigm cases? What choice was made, how was the dilemma resolved, what safeguards were introduced, how was the decision justified? Then ask, what happened? What was the outcome? Who benefited, and how? Who got hurt, and how? How did the public/media/regulators react to the choices made? How did they respond to the justifications given? Did those who made those choices come to regret them, or renounce them, or were they openly proud to have made them? Did this case provide a template or model of ethical success, or does it function as a warning?

4. **Use Analogical Reasoning to Identify Parallel Risks, Opportunities, Solutions, Risk Mitigation Strategies**

   This is the tricky part. Knowing how the paradigm cases both do and don't resemble this one, and considering the uncertainties involved (history does not *always* repeat), how should our ethical knowledge of the paradigm cases influence our ethical reasoning and judgment in *this* case? What lessons should transfer over? What solutions that worked well before are likely to work well again? What mistakes that they made then are *we* in danger of making *now*? What risks that were successfully mitigated *that* time can be mitigated with similar strategies now? And how might the relevant differences between the cases limit or alter the transferability of the paradigm lessons to this present case?

Implementation Case: Company D is designing an AI virtual agent (AIVA); the target audience is corporate executives who want to reduce their reliance on human assistants in the office. Early on in the design process, the team sits down to do a case-based analysis of the risks, including ethical risks, of the virtual assistant. They begin by analyzing two examples of AIVA's that were brought to market successfully, a case in which the AIVA never made it to market, and a case in which AIVA went to market but failed. Most of the paradigm cases involved some number of ethical concerns; the successful ones mostly were able to address them, although in one case the product remains a commercial success but is the subject of growing media and regulatory criticism.

The team starts by identifying the parallels and dissimilarities. All the cases involve some level of privacy concerns, including the present case. Several of the cases raised some ethical concern about replacing human workers, as does the present case. Two of the paradigm cases were AIVA's built to interact with children, and two were for general use. The present case, however, targets a narrower and more socially powerful userbase: corporate executives. So, some of the ethical issues presented by AIVA's interaction with children do not apply here. But corporate executives handle much more legally and economically sensitive data and transactions than most other adults, so the privacy and security issues here are recognized as even more acute. All the cases involved public controversy about the female gender presentation of the AIVAs; given the under-representation of women in corporate management, and the old stereotype of the female secretary, this ethical issue is seen as even more sensitive here than in the general cases.

The team then looks at how the designers/developers in each of the other AIVA cases chose to handle the ethical risks, tradeoffs, and challenges involved in their project (so far as those choices can be inferred from the available information.) They take note of the outcomes in each case, and where the result differed from what was probably desired or expected. They look at which risks/worries in the previous cases, such as worries about disastrous impacts on human workers, turned out to not be a big problem (in fact, the labor impact was marginal in all the paradigm cases); but they also look at whether this case suggests any different trajectory.

They use analogical reasoning to transfer over design solutions, trade-off compromises, and risk mitigation strategies that worked well and seem like they should work again; in other cases they decide that the present case is unique in its ethical circumstances and requires a new solution. They also develop a set of most likely scenarios/outcomes for the present case and an ethical response plan for each one, so that they are prepared to react quickly and wisely to plausible outcomes, beyond the outcome that they most expect or desire. In fact, many of their transferred solutions/strategies work as hoped, but one solution backfires; fortunately that ethical failure is covered in one of their planning scenarios, and they have an intelligent and well-framed response ready to address it. The company's prompt and sensitive response to the problem saves the product from a PR disaster and commercial and ethical failure—all for the modest cost of a planned team exercise in ethical design case-based analysis.

TOOL 5: REMEMBERING THE ETHICAL BENEFITS OF CREATIVE WORK

It is important, as we have seen, for designers and engineers to focus on ethical risks. But if we aren't careful, this can lead us into forgetting that ethics is about a *positive* outcome; it's about human flourishing, including that of future generations, and the promotion of healthy and sustainable life on this planet. Great creative work advances those aims, and ethical design and engineering is a powerful form of such work.

Yet sometimes, the short-term and less ethically-grounded benefits of our work (the raise, the good performance review, the praise of our boss or the board, the smooth investor call, the quarterly bonus, the stock jump) eclipse the greater goal that motivated us to do this creative work in the first place.

In the worst-case scenarios, the loss or perversion of ethical motivation leads to massive corporate corruption and failure (e.g. Theranos, Enron) or to a disaster that the company manages to survive, but with a damaged reputation and lost competitive advantage (Uber), and/or with people going to jail (Volkswagen). In other scenarios the damage may be more subtle, but no less real—a tarnish on the brand, slowed growth, failure to innovate in ways that people care about, departure or demoralization of those talented individuals who want an ethically rewarding work environment, recruiting failures, and growing employee apathy, depression, anxiety, detachment, or cynicism.

To counter this, it helps to implement a workflow tool that makes those ethical benefits explicit and deepens sincere motivation to create them. It is important that this exercise not devolve into patting each other on the back and self-congratulatory praise for 'making the world a better place'—it is *hard* to accomplish that and it should always be framed as the goal we work towards, *not* the thing we smugly celebrate ourselves for having done, or the thing that we believe we are 'destined' to do because of our smarts or our goodness.

To keep the ethical benefits of creative work at the center of the team's or the company's motivational set, find ways to together ask hard questions like these:

- *Why* are we doing this, and for what good *ends*?
- Will society/the world/our customers *really* be better off with this tech than without it? Or are we trying to generate inauthentic needs or manufactured desires, simply to justify a new thing to sell?
- Has the ethical benefit of this technology remained at the *center* of our work and thinking?
- What are we willing to sacrifice to do this *right*?

Implementation Example: The leadership of company E notices that morale company-wide seems to be sagging despite strong corporate profits. Internal data suggests that employees are becoming more cynical about the company's values and impact on the world, more 'checked-out' from the long-term vision of the company, and more focused on just 'getting paid and moving on to a better opportunity.' Recruiters report difficulty securing some of their best prospects, who seem to be worried that this company is 'just about the stock price now' and not really invested in making positive change any more.

The leadership organizes an all-hands meeting dedicated to explicitly revitalizing the ethical culture of the company, and reaffirming the company's dedication to the ethical benefits of its work. Sensitive to the likelihood that cynical employees may see this as a

self-serving or pointless exercise, the leadership seeks input from highly respected employees of all ranks on how the message can be delivered and made sincere. They conclude that some key policy changes are needed in order to demonstrate the company's sincere interest in revitalizing the ethical mission, so they implement and announce those changes prior to the meeting. They also develop an anonymized survey instrument, administered by a trusted third-party, which will ask employees to answer the questions above, and to offer input as to what further changes, if any, would strengthen the company's ethical mission and resolve. Part of the all-hands meeting involves examining and openly discussing that anonymous feedback.

TOOL 6: THINK ABOUT THE TERRIBLE PEOPLE

When former Google CEO and Alphabet's chairman Eric Schmidt spoke at the RSA Conference in San Francisco in 2017, he said the following: "We now find ourselves back fixing [the Internet] over and over again," Schmidt said. "You keep saying, 'Why didn't we think about this?' Well the answer is, it didn't occur to us that there were criminals."

Positive thinking about our work, as we saw in Tool 5, is an important part of ethical practice. But sometimes what can be a virtue becomes a vice, as it does when we imagine our work being used only by the wisest and best people, in the wisest and best ways.

In reality, technology is power, and there will always be those who wish to use that power in ways that benefit themselves at the expense of others. And there will be those who use the power we give them for no rational purpose at all. If you are building or granting access to powerful things, however, it is your responsibility to *mitigate* their abuse to a reasonable extent. You don't hand a young child a kitchen knife or lit candle and walk away. You don't lend an arsonist your canister of gasoline and a match. And you don't make an app that collects health data and transmits or stores it with weak encryption.

So, these questions need to be asked at key design stages:

- Who will want to abuse, steal, misinterpret, hack, destroy, or weaponize what we built?
- Who will use it with alarming stupidity/irrationality?
- What rewards/incentives/openings has our design inadvertently created for those people?
- How can we *remove* those rewards/incentives?

Implementation Case: Company F is known for its baby monitors and home security devices. Its designers come up with an idea for a device that will allow parents, babysitters, and daycare workers to have real-time locational monitoring of the children

under their care, through an app that interfaces with tiny RFID tracking devices embedded in their children's shirt collars. The collar device can also be activated with a hard press by the child to send an 'alert' signal to the app, which will text and notify the parent or sitter to check on their child. However, early on in the design process, the designers go through a mandatory 'think about the terrible people' exercise in which they envision all of the ways in which this technology is likely to be abused or misused.

The design team quickly realizes that although they initially envisioned use cases like caregivers watching TV in the living room or bathing an elderly parent while allowing young children to play safely upstairs or out in the yard, in fact some neglectful caregivers will be tempted to use it to justify leaving young children at home alone (since they will be notified by text if the children leave the home perimeter, or call for help). They also realize that the collars will continue emitting a locational tracking signal beyond the home perimeter, so this information could be used by networks of child predators looking for a visual map of unattended children in the neighborhood. They also realize that nothing prevents the device from being surreptitiously fitted into the collars of teenagers, spouses, partners, workers, friends, or enemies. As they consider all of the possible unethical use cases, they realize that their initial design was profoundly flawed and unsafe; they resolve to abandon this particular design project until they can make the application more readily restricted to appropriate use cases.

TOOL 7: CLOSING THE LOOP: ETHICAL FEEDBACK AND ITERATION

The *most* accurate and helpful way to think of design/engineering ethics is as an ongoing *process*; the *least* accurate and helpful way is to frame it as a task to be completed. Ethical design is a never-ending loop that we must ensure gets closed, to enable iteration.

Now, ethics is *not* relative in the way that egoists and sociopaths might understand it to be (where ethics is just 'whatever those with power say it is'). Human flourishing and the sustainability of life on this planet are always ethical goals, and nothing that makes those goals impossible, or harder to achieve overall, can be ethical (and no, those goals are *not* incompatible). But ethics *is* relative to the particular social context in which those goals are sought; a technology that promotes human flourishing in one social context (the controlled use of narcotic painkillers under close medical supervision) can undermine it in another social context (an unfettered market of 'pill farms' and doctors paid by pharmaceutical companies to overprescribe narcotics and ignore abuse).

Because society is always changing, and we with it, the ethical impact of technology is always a moving target. This is even more so because technology *itself* continually reshapes the social context, and the people in it. A device or piece of software whose design is robustly ethical on its launch date may be *un*ethical two years later, if the social conditions and user base have changed enough that its impact is no longer conductive to human flourishing.

This means that ethical reflection, analysis, and judgment are perpetual elements of good design and engineering; these skills never stop being needed.

Thus it is important above all to bring design and engineering culture—especially in the tech industry, where the professional safety culture of civil and mechanical engineering is not well-embedded—into the domain of ethical practice as a *permanent* shift. Ethics is not an external requirement or addition to good design and engineering. It is not something to be 'checked off' and forgotten. It is a way of staying anchored in the best possibilities of our chosen profession, a way of becoming and remaining the kind of designers and engineers we want to be.

To embed this understanding in a company culture, some concrete steps are needed:

1. **Remember That Ethical Design/Engineering Is Never a Finished Task**

   To make this concrete, ethics needs to become part of institutional and team memory. In addition to being an explicit part of what the team is doing now, it needs to become part of how we describe what we did in the *past*, and why we succeeded in the fullest sense of *good* technology design and engineering—not just as a commercial success. It also needs to be part of how we talk about the *future*; the kinds of design and engineering success we *want* to have and *will* have through sustained ethical practice. When ethics is presented as part of the organization's past, present, *and* future, it takes its proper place in the company's culture and identity.

2. **Identify Feedback Channels that Will Deliver Reliable Data on *Ethical* Impact**

   There's no way to know whether, or to what extent, our work is actually succeeding in that fullest sense unless we are gathering reliable data on the *ethical* impact of our designs on society, and on specific stakeholders. That kind of data does not come unless instruments are designed specifically to elicit and transmit feedback of that kind. Any product design plan should identify specific instruments/modes/channels by which ethical impact data will be collected; from users, certainly, but also from other affected stakeholders and groups. Likely impacts on ethically important institutions (democracy, education, media), cultural elements (art, literature) or physical systems (the food chain, oceans, climate), which as non-persons cannot speak for themselves, must also be audited in some fashion.

3. ***3.*** **Integrate the Process w/Quality Management & User Support; Make it *Standard***

   The auditing of ethical impact cannot be an *ad hoc* event; it must become a standard feature of product quality management. Wherever possible, it should be integrated with QA/QC and user support processes that are already standardized, without compromising the ethical sensitivity of the audit instrument (for example, 'high user engagement' is not itself an ethically sensitive metric and must not be taken as such).

4. Develop Formal Procedures and Chains of Responsibility for Ethical Iteration

How will ethical audit feedback get *analyzed*, *communicated*, and *used* in the next design cycle? Who will be *accountable* for closing the loop? Who is accountable for ethical design *overall*? These things will not happen on their own. Left to the 'good intentions' of ethical people who are nevertheless incentivized by the company to prioritize *non*-ethical metrics of success, a company cannot claim to be surprised when ethical disasters or decline are the outcome. A company gets whatever its incentive structures and chains of responsibility reward. If done right, formal procedures and chains of responsibility for ethical design and engineering do not make ethics 'impersonal' or 'inauthentic'—they make it *effective* and *realized in the world.* A company with a 'values statement,' a 'Chief Ethics Officer,' and a mandatory ethics onboarding training, but which *does nothing to formally operationalize and incentivize ethical practice,* is relying on little more than ethics vaporware.

Implementation Example: Company G is committed to instituting a robust culture of ethical design and engineering, in ways that will make its long-term success more sustainable and that will help to earn back the eroding public trust in technology's promise for humanity. To do so, they make sure to set up the necessary structures to formalize and incentivize ethical practice, and to 'close the loop' of ethical design and engineering so that the company remains attentive, agile, and responsive to a constantly evolving social context that can reshape the ethical landscape for their products in unexpected ways. They integrate ethically-laden language in the company's articulations of its past and its future. They take existing structures and channels for product quality management and user support and enhance them to elicit ethically relevant feedback about the impact of their products on the flourishing of diverse individuals, groups, institutions, cultures, and systems. They create formal procedures and responsibility chains for ethical design and engineering, including implementing many of the tools in this toolkit into existing design and engineering workflows.

These measures to 'close the loop' enable continuing ethical refinement of their products, and help to ensure that urgent ethical problems are addressed quickly and adequately. Later, they learn from one of the ethical feedback channels that their newest product is being abused to enable targeted violence against a particular ethnic minority in a remote region. The urgent alert was submitted via that channel by local NGOs, and promptly forwarded to a manager in the QC division who is trained and empowered to organize a rapid company response to just these kinds of ethical issues. After convening the appropriate technical and managerial team, a higher-level VP for ethical product design makes the decision to instruct the team to rapidly push a product update that temporarily suspends the specific functionality being exploited in that region, and communicate this solution to the NGOs, requesting local confirmation of its efficacy. A reasoned company response explaining the change and its consistency with its ethical principles and values statement is circulated internally, along with explicit credit and thanks from the CEO to the QC manager, VP, and others on the team who led the response.