

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Mechanism

Creativity itself is value-neutral. Works that are new and useful can have positive, negative, or neutral effects. Sometimes all three, depending on the situation or the people who use them.

Ethics brings more clarity in determining positive and negative effects. It is concerned with «What is good/right vs. what is bad/wrong?», including prescriptive assertions about what should be done. Some of it is implemented in law, though much of it remains societal expectation or aspiration.

Ethics is not monolithic. There are different theories of ethics, e.g., Utilitarianism, Virtue Ethics, Deontological Ethics, and many more. In creativity, especially in product development, there is often a strong focus on Utilitarianism. While it is important that products work, there are also red lines that must not be crossed (Deontological Ethics). Thus, combining different ethics is usually recommended.

Applicability

Ethics is relevant whenever you create something — no matter how large or small, or whether for others or yourself. The strength

may vary. Few will reach Frankl's warning of «... let us be alert — alert in a twofold sense: Since Auschwitz we know what man is capable of. And since Hiroshima we know what is at stake.» (see also Box 1: Pressure on Ethics). That is not an argument against ethics in «ordinary creative projects», it just defines the scope — whatever you create can have positive and/or negative effects.

Given that creativity is value-neutral, the work itself will not ensure that it is used ethically. You can go wrong in both directions. Work that might appear unethical can be used for positive purposes, e.g., work on learned helplessness can be used to develop interventions that preserve and restore agency. Similarly, work that appears beneficial can turn into nightmares, e.g., a smart city can become a digital concentration camp, or gamification for hotel cleaning an electronic whip.

Ethics also applies to art. Art can inspire and rouse emotions — just look at the ways art is used for propaganda or advertising. While scientifically questionable, some artworks have been blamed for negative effects such as suicides («Werther effect», though more likely an early example of a moral panic). There are also ways to use one's creativity against the dumbing of the public. Adbusters, for example, creates campaigns against advertising.

Ethics strongly affects science and engineering, especially digital product development. Typical issues are, among others, access to data (user profiles, prediction, influence), information flow (manipulation, filter bubbles, censorship), attention economy (see also Digital Environment), and control/pater-

nalism (persuasive technology, gamification, nudging). Additionally, research and product development usually deal with target audiences or participants, which come with their own ethical challenges (see Target Audience Feedback).

Ethics needs to be applied especially when it is inconvenient. When things are fine, few people have a problem with ethics and almost everyone agrees that it is important. The actual test of ethical behavior is when things are tense — for example, when money, health, or organizational or personal survival are on the line (see also Legitimacy and Corruption). Unfortunately, good intentions are no defense. On the contrary, they are a warning signal (see Box 2: Why Good Intentions are a Warning Signal).

Ethics has become a larger issue today, in part due to ethical failures of the past (see Box 3: Ethical Guidelines and Box 4: Ethics Commissions), as well as increasingly paternalistic governments and larger bureaucracies. In some cases, ethics is used as a cudgel for ideological or partisan issues. In these cases, the pendulum has swung too far into restriction.

Given that ethics can be used as a weapon to enforce a partisan agenda or make people feel good while preventing new and useful work, knowing about ethical issues can help you defend against these attacks. You can reduce the attack surface in advance and counter challenges.

Thus ethics affects all creative projects:

- I am making something for myself: Ethical Standards

Box 1: Pressure on Ethics

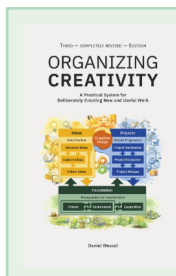
«There is an old Rabbinical story that's passed down through the ages. And it goes like this: <The people in Noah's day were very wicked, which is why God send the flood. And the people who didn't make it upon the boat, which were many — when the waters rose to their knees, they pulled their children to their waist. When the waters rose to their waist, they put their children in their arms. When the waters rose to their neck, they put their children on their shoulders. But when it rose even further, they put their children under their feet so they could survive.> When the threat of mortality comes down from so many avenues, in over every medium, over every channel, everyone is telling you, [...] you're going to die if you don't take these interventions that we recommend — it does crazy things to society.»

Justin Hart

- I am releasing work to users/audience: Creative Works
- I am testing with people/animals: Target Audience Feedback
- I collect data or influence behavior: Creative Works + Target Audience Feedback
- I face pressure to cut corners: Ethical Standards + Watchdogs
- I suspect harm after release: Remediation / Educate / Better Version / Whistleblowing

Relevant Chapters

For background information, see Chapter 1: Creativity, 12: Creative Commitment, and 14: Project Evaluation.



DRAFT VERSION FOR FEEDBACK

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 2: Why Good Intentions are a Warning Signal

One of the common defenses against unethical behavior, especially the unintended consequences of one's actions, is «but my intentions were good».

That is usually the problem, not a valid defense.

The moment a person says «I have good intentions.» is usually the moment they realize that they might violate an ethical boundary. Otherwise, they would not need to refer to good intentions. It also encourages the continuing thought of «So what I do must be good as well.» Thus, it encourages a person to miss or downplay negative effects. Sometimes those negative effects are blamed on external circumstances or other people, e.g., bad luck, sabotage, stupidity, the user (PEBKAC, DAU), etc. Anyone but the person who created it, because the intentions were good, after all.

Good intentions are also extremely common, even in the worst people. You can be pretty sure that even the worst fascist or communist butchers claimed they had «good intentions». It is a rationalization, not a safeguard.

An even worse form is when «good intentions» are equated with «we are the «good ones»». Once «being good» becomes an identity marker, the «good» person's actions automatically become «good» and the person is effectively immunized against criticism when they commit harm. They cannot be responsible for the negative effects; others have to be responsible. Very comforting, but besides harming others, it is not helpful for creating something new and useful.

In this sense, good intentions are understandable, but not good. If you refer to «good intentions», take it as a warning signal to become very, very careful in what you are doing. There is a reason that the road to hell is paved with them. Instead of relying on them, test the consequences of your actions with the same rigor you would test a competitor's work.

Goodness must be tested, not presumed. Good intentions are not sterile — they contaminate decisions like anything else when you don't examine them. See also Box 17: Good Intentions Quotations on page 15.

Box 3: Ethical Guidelines

Some disciplines have ethical guidelines or standards to which practitioners are expected or required to adhere. They have a long history (e.g., the ancient Greek «*primum non nocere*») = «first do no harm», which continues with being careful and healing — not the worst guideline).

However, these guidelines are no guarantee that they are actually ethical or interpreted in an ethical manner. They are written by institutions that can be co-opted by activists and other ideologues. As those people pursue a political or ideological agenda, they are more interested in what is useful to their goals. In science, activism can easily override scientific integrity — when the results and the ideology clash, ideology usually wins via suppression of inconvenient truths (see Box 5: Ideological Capture).

A related problem is that ethical language is often rather abstract — it has to be, otherwise it would not be widely applicable. But that abstractness creates too much room for interpretation. What does «dignity», «harm», «common good», or «care» actually mean? For example, «care» can mean assisting people who have lost agency to regain it, or acting paternalistically and removing that agency by saying these people

need others to care for them. There is a lot of power in defining the terms. At their worst, ethical guidelines can be perverted simply by redefining these terms.

A well-formulated example of an ethical guideline is the «Code of Ethics» of the American Psychological Association (APA, 2017). It has the principles «Beneficence and Nonmaleficence», «Fidelity and Responsibility», «Integrity», «Justice», and «Respect for People's Rights and Dignity». It then applies these principles to concrete standards, e.g., informed consent, which is defined as providing information, understanding that information, and voluntary participation. Still, it does not take much to pervert these codes, e.g., by defining beneficence on the group level and thus overriding the individual. Or by simply putting a group of people outside the category «human» to justify any treatment — the German «Untermensch» comes to mind. Unfortunately, humans are — and were — very creative here.

Thus, while ethical guidelines can provide orientation, one should never delegate one's conscience or moral judgment to them. In moral panics they can also be put aside by those in authority. Acting ethically means doing what is right, regardless of what you are told (see also □ Saying No).

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 4: Ethics Commissions

Ethics commissions, also called ethics committees, institutional review boards (IRB), independent ethics committees (IEC), ethical review boards (ERB), or research ethics boards (REB), are supposed to review planned scientific research and point out ethical issues.

At their best, they keep science ethical and honest without making it unnecessarily difficult. At their worst, they either block research on ideological grounds (see also Box 5: Ideological Capture) or succumb to the fox guarding the henhouse.

There can also be confusion when ethics commissions are rooted in a particular discipline but decide on research in other areas. For example, in medicine the focus is the human body — the issue is health and once a treatment was given, you cannot undo it. In psychology, the focus is the human mind — while studies can have strong emotional effects (e.g., Milgram experiments, Stanford Prison study), people can usually remove themselves and leave. The ability to stop participation without negative consequences is one of the cornerstones of the ethical code for psychologists. In design, the focus is an artifact. Users do not fail; the artifact did. Apps are evaluated, not the users' intelligence (mind), or the effects of a drug on the human body. Using an ethics commission rooted in one discipline, e.g., medicine, to judge ethics in another, e.g., psychology or design, is a mismatch.

Traditional ethics commissions also likely

have blind spots regarding the development of creative works:

- **Commercial projects can have conflicts of interest:** The goal is often not to find out what is true, but to create a successful product. That introduces incentives usually not present in research, though it can be an issue with, e.g., medical products.
- **Co-creation conflicts with anonymity:** Participants can contribute with their creativity in co-creation workshops, which deserves public recognition (also in the minds of some participants). This conflicts with the usually expected anonymity in research.
- **Casual data/covert observations:** Digital products allow the collection of lots of casual data. This is usually not available in classic lab research.
- **Legal issues regarding terms of service:** Research using existing services can conflict with their terms of service (e.g., crawling an online dating platform).
- **Values embedded in artifacts:** Some products are deliberately developed to affect social change (e.g., climate change, eating meat, etc.), but any product has affordances even without an explicit behavior-change intervention. Informed consent addresses participation in the study itself, but what if the participant does not agree with the overall goal of a product, yet data based on this participation in the study has made it better?
- **Nudging, gamification, persuasive technology:** Sometimes values are not only embedded, but behavior change is the goal. Are techniques such as nudging, gamification, or persuasive technology ethical? Or are only dark patterns too much?
- **Societal effects of products/changing socio-technical system:** What about the other effects of the product? For example, side effects and long-term effects, different use (cf. security mindset), safety-critical products, predictions/social influence (individualized advertisement, election interference, etc.), change of information flow (filter bubbles, censorship, paternalism), control of attention (user is the product for social media; if the user is offline, the company makes no money), influence via changed reward structures, mob behavior (e.g., Twitter storms). They are hard to assess, so an ethics commission will likely miss them as well.
- **Wasting resources:** Is resource use something that should be evaluated? For example, a single-use pregnancy test with the hardware to run «Doom» (IRBs can see «research waste», but can they assess waste in products?).
- **Artificial intelligence:** This is a topic with lots of new ethical issues in itself (e.g., fairness, autonomy/control, transparency, dependability, safety, security) and long-term issues due to model drift, skill degradation, etc.
- **Ethical regulations due to profession:** What if the domain comes with its own ethical guidelines? For example, the German society for Computer Science

has its own criteria for professional ethics (different kinds of competencies, good judgment, work conditions, courage, social responsibility, etc.), including when it comes to surveillance technology.

- **Community rules (developer/hacker ethics):** What about more bottom-up, community-driven rules of good behavior, including the use of technology, freedom of information, rules for writing code, etc.?

These issues do not mean that an ethics commission should not be used. But they have blind spots — neither a positive nor a negative vote is evidence in itself.

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 5: Ideological Capture

Ideological capture occurs when a discipline, organization, or project subordinates its own telos to an ideological frame. All people have political, moral, religious, or philosophical commitments. The problem starts when those commitments determine in advance what may be asked, what may be found, what may be made, who may speak, and how dissent is treated.

Here, ideology means a closed explanatory and moral system that treats its own premises as unimpeachable and subordinates truth, craft, beauty, usefulness, and responsibility to its own preservation. This often happens without bad intention. Its adherents may simply assume that the ideology is true, that they are «on the right side of history», or that they are doing good. The danger is precisely that good intentions make the closure feel justified (see also Box 2). The intention may be good. The effects are not.

In science, capture replaces truth-seeking with moral-political acceptability. In art, it replaces beauty, craft, perception, or expression with ideological service. In organizations, it replaces the stated purpose with loyalty to a cause, faction, or identity. The ideology may be left-wing, right-wing, religious, nationalist, corporate, bureaucratic, or activist.

In many current Western institutions, these mechanisms often appear in left-coded forms — diversity language replacing merit, care language replacing agency, harm language replacing evidence, or inclusion language

justifying exclusion. In other institutions the same mechanism may appear as nationalism, religious conformity, corporate loyalty, security logic, or party discipline. The diagnostic question is not which side uses the language, but whether the work's own standard has been replaced by ideological loyalty.

Telos Gets Replaced

- **Mission drift:** Issues only loosely related to the discipline become central, while the discipline's own standard is demoted. Science becomes moral positioning instead of truth-seeking. Art becomes message delivery instead of artistic work. Engineering becomes symbolic compliance instead of working solutions.
- **Identity over work:** The creator's demographic, political, religious, or institutional identity matters more than the quality of the work. A useful test: Would the work be judged differently in a blind assessment?
- **Moral goal over method:** A noble goal — saving the world, protecting the environment, ending discrimination, defending tradition, protecting the nation, serving God, maximizing shareholder value — is used to excuse weak evidence, poor craft, coercive methods, or selective standards.
- **Consensus over evidence:** Consensus is treated as if it settles the question, but it can only summarize the current state of a field, not determine what is true. It cannot replace evidence, method, uncertainty, replication, or the right to challenge the claim.

Political or Institutional Pressure

- **Power transfer:** Science, art, or ethics are used to legitimize decisions that are actually political. Even the best evidence can only inform action; it cannot by itself decide what should be done. That requires values, trade-offs (compared to what? at what cost?), and responsibility.
- **Dependency pressure:** Funding, publication, promotion, exhibition, access, reputation, or employment depend on affirming the preferred frame.
- **Selective standards:** Rules are applied strictly to opponents and loosely to allies. The standard changes with the target (e.g., selective rigor).
- **Compelled affirmation:** People are not merely asked to comply with lawful or professional conduct; they are expected to affirm the ideology's premises, e.g., diversity statements, loyalty oaths, mandatory political declarations, public statements about current issues, or required affirmation of institutional values beyond professional conduct.

Moralization

- **Question becomes accusation:** Asking whether something is true, effective, beautiful, legal, useful, or well-made is treated as evidence of bad character.
- **Harm language replaces argument:** Claims are rejected because they are called harmful, unsafe, offensive, or dangerous, without showing that the claim is false or the work is defective.

- **Dissent becomes moral contamination:** People who disagree are not treated as wrong, mistaken, uninformed, or differently situated. They are treated as evil, dangerous, contaminated, or illegitimate. This functions as a discourse-stopper: the claim no longer has to be answered and the ideology presents itself as identical with goodness itself.
- **Responsibility is delegated:** People are asked to outsource their judgment to the ideology, its representatives, or its approved vocabulary.

Taboos and Fear Signals

- **Topics become unspeakable:** People lower their voice, avoid certain questions, or continue conversations only in private.
- **Humor disappears:** The ideology treats itself as sacred. Jokes, distance, irony, are treated as threats rather than relief valves.
- **Self-censorship becomes normal:** People know what they are not supposed to notice, mention, or create.
- **Exit costs rise:** Disagreeing threatens career, access, reputation, friendship, publication, funding, or safety.

A single red flag does not prove capture — the pattern matters. When ideology becomes sovereign, truth is useful only when it confirms the frame, beauty only when it serves the message, craft only when it advances the cause, and ethics only when it protects the ideology's legitimacy. Creativity is not killed all at once; it atrophies into the performance of loyalty.

Ethics in Creativity Worksheet

Intervention Variables

The intervention variables are divided into Ethical Standards, Creative Works, and Target Audience Feedback.

Ethical Standards

It is hard to define universal standards for ethical behavior, though the following aspects are useful:

- **Responsibility for one's actions:** Moral judgment cannot be delegated, though people often try to convince others or themselves that it is possible. Others to make them commit atrocities («You have to follow orders.»), themselves to wash their hands of what they did («I was only following orders.»). Taking responsibility means that you decide what is ethical and what is not. External checks, affected-user perspectives, law, domain standards, and consequences provide information, but the responsibility is yours.
- **A combination of deontological and utilitarian ethics:** Each on its own would be insufficient.

Deontological Ethics: There are red lines that must not be crossed. Kant's «So act that you use humanity, in your own person as well as in the person of any other, always at the same time as an end, never merely as a means.» is a good line.

Utilitarian Ethics: Principles are only decoration if they do not have the desired effect. Thus, the behavior must also improve the situation, or at least not make it worse.

- **Craftsmanship:** The works you create actually work, are in good condition, and do what they should do. What they do and cannot do is communicated clearly.
- **Legitimacy:** What makes a decision or company legitimate? Necessity explains only motive, urgency or risk explains pressure, and panic explains overrides — but none of them are legitimate. They are the situations against which ethical principles should protect. For example, in health emergencies, ethical principles should constrain options, not be thrown out because the situation is «different» or «too urgent». In many situations, only the informed consent of a person affected by the decision makes it legitimate.
- **Methods over Goals:** A common failure is thinking that noble goals justify any means, especially when that goal is a utopia that can never be reached. Then the abstract utopia creates a concrete hell. Instead, the process has to be ethical, even if you disagree with the outcome. Unfortunately, many people are willing to accept unethical methods (coercion, physical force, etc.) if they agree with the attempted outcome. That the same methods will then be used against them once an outcome is pursued with which they disagree is outside the scope of their imagination.
- **Ethics is steady, but not loud:** People have an innate need to appear moral. Given that immoral behavior can lead to ostracism and ostracism kills, that makes complete sense. However, it can tempt people to become self-righteous. Thus, ethical behavior is steady, but not loud. It has no need to prove itself or be admired.
- **Look out for side-/long-term effects:** Life is a system. Whatever you do has unexpected effects and some of them only become visible over the long term. Unless you look out for them, they are easy to miss (see also □ Diagnosis).
- **Watch the Trajectory:** Problems often announce themselves if you are vigilant. For example, growth numbers might not be sustainable, so despite things looking good, you and your employees may end up in a bad situation soon. Intervening early can allow you to change the project while you still have resources (see also «Quit» by Annie Duke).
- **Low Waste:** Protecting the environment should not become politicized or co-opted by ideology. Pollution is waste, and waste is expensive. Thus reducing waste is useful for almost everyone except waste-disposal companies and activists. Even a little waste scales quickly if the product becomes a success, so mind the small stuff and show everyday responsibility to protect the environment. It conveys the right message and helps ensure the right mindset if larger issues appear.
- **Get Watchdogs:** Creative people are good at coming up with justifications, including for their own behavior. Getting feedback from independent, knowledgeable people helps you stay honest by pointing out things you do not see. This can be an ethics commission or an external consultant — as long as the fox is not guarding the henhouse and they face no negative

DRAFT VERSION FOR FEEDBACK

consequences for telling you the truth. Even when they advise you to kill your baby.

«To paraphrase Bill Walsh, when left to our own devices, many of us individuals seek lower ground like water. The key, then, is to support our natural inclination to justice with strong boundaries and strong commitments—to embrace, as Lincoln urged a divided, angry nation to do, «the better angels of our nature.»

Holiday & Hanselman (2016)

- **Educate the Public:** If your work has become harmful and is outside your control, you can educate the public. If it costs money — and loss of reputation often does — people are more prone to listen. Use your expertise and learn from influence groups that do not discredit themselves with ideology (far left/right) or violence. Contact groups that can help you (e.g., environmental groups), but be careful whom you choose as allies. Other groups often have their own agenda and methods, and even if their aims match yours well enough, their methods might discredit or even criminalize you. Note that even as inventor, main contributor, or expert, success is not guaranteed. If you go against the mainstream, especially in a moral panic, you will likely be ostracized by mainstream media. One effective option is to show the mechanisms, e.g., how attention economy manipulates the user, or how dark patterns are used.
- **Create a better version:** You were creative enough to bring it into the world; who says you cannot try to stop it or improve it

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

to a point where the dangers vanish? Find ways that do not make matters worse to stop the danger, e.g., develop a better process to reach the same goals without the side effects.

- **Blow the whistle if necessary:** In extreme cases, e.g., Fraud, whistleblowing might be the last option (see Box 6: Fraud). As this usually has negative consequences, it is a last-resort option (see □ Whistleblowing).

Box 6: Fraud

Fraud is a common ethical violation. For example, commercial companies may sell products without the advertised effects or hide negative effects. Art can involve forgeries or be used to launder money. Science has plagiarism, data manipulation, and fabrication. The question is not why fraud occurs, but why it should not occur.

A useful model is the «fraud triangle» by Cressey. Fraud occurs when motive/pressure, opportunity, and rationalization come together. For example, regarding fraud in science:

- **Motive/Pressure:** Publications need «good» data, and publications are required for a career in science.
- **Opportunity:** Often only the scientist has access to the data, so manipulating the data is easy.
- **Rationalization:** The theory is seen as correct; only the data is dirty. In reality, the data would look better.

Creative Works

When it comes to creative works, the following aspects can be relevant (the term user applies to all target audiences, including readers of a book or people looking at artworks):

- **Avoid Creative Waste:** Is this use of attention, money, material, data, or audience trust justified by the likely value? It is easy to create works that offer nothing new and useful, or are just tiny variations.

As the pressure will continue to exist (competition in science and art), and creative people are usually creative and intelligent enough to rationalize anything, reducing the opportunity to commit fraud is the best way to reduce it. For example, in science this means ensuring that data are collected and stored by multiple people, including those without vested interests.

Fraud is not caused by a lack of ethical training or knowledge, nor can it be prevented by formal authority. Stapel, who fabricated scientific findings, was teaching his department's scientific ethics course. Knowledge does not prevent misconduct, sometimes it gives the person better tools.

«It is apparent from Mr Stapel's statements [...] that the questionnaires and additional materials loaded into the back of his car (e.g. bars of chocolate for use in an experiment) never went to schools but were dumped in a container. Mr Stapel himself would then sit at his computer with an empty questionnaire to create a corresponding dataset.»

Levelt, Noorth, & Drenth Committees (2012)

Address issues that are meaningful, that matter, and where your solution can actually help people or «push the human race forward». If you are doing something that is not worthwhile, creating something each and every day that does not move your heart, then perhaps it is time to stop, think, and find another solution.

- **The Golden Rule is not so Solid:** Whether it is Kant's «Act only in accordance with that maxim through which you can at the same time will that it become a universal law.» or «Do not impose on others what you do not wish for yourself.» or «Therefore all things whatsoever would that men should do to you, do ye even so to them.», the golden rule has a huge blind spot. People's tastes differ. For some people, needles are a nightmare; for others, pleasure (needle play). Going by what you would not want done works for you, but not necessarily for others. So the user must always have the final word. Otherwise, care is just coercion in a better suit.
- **The user is never merely a means to an end:** Kant's formulation works here. While making money is what keeps most creative companies running, it must never reduce the user to a means to an end, irrespective of how good that end might appear. This includes noble goals such as saving the world. The dignity of the target audience is inviolable.
- **Self-Determination:** Using self-determination theory by Deci and Ryan (e.g., 2000), humans have the following three universal basic needs:

Autonomy: The user has agency, decides for himself which interests to follow, has actual choices («Whoever controls the options has the power.»). The user also decides whether to use the work in the first place, when to stop, and whether to continue using it. This prohibits pressure, coercion, and interruptions or distractions if not absolutely necessary.

Competence: The capabilities of the user are supported and enlarged, learning how to use it is easy, and his self-efficacy is improved. Ideally, a work should increase competence and then no longer be necessary. This prohibits making the user dependent on it, skill degradation, or overwhelming him.

Relatedness: The user is assisted in meaning-making, building significant relationships, and can contribute and receive support from others if relevant. This prohibits using popularity or status as extrinsic motivation.

- **Informed Use:** Never deceive the target audience about the actual goal. If the work aims to persuade the user, be open about it. Do not deceive.
- **Opt-In as Default:** Users decide what they use and share.
- **Data Protection:** Users need to be informed in a quick and easy-to-understand manner what happens and what can happen to their data. Even if you ensure confidentiality and keep it, can you protect them from government searches?
- **Accountability:** Many products work better if users are accountable for their

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

actions. This does not mean that they are identifiable — anonymity has its advantages. But the user's actions should have consequences attached to that user. Nicely put in Wired: «The problem isn't anonymity; it's accountability. If someone isn't accountable, then knowing his name doesn't help. If you have someone who is completely anonymous, yet just as completely accountable, then — heck, just call him Fred.» (Schneier, 2006).

- **Get Watchdogs:** If you created something, it is hard to imagine that «your baby» could do harm or be used for bad. Get external feedback, ideally from people with a security mindset («looking at the world in terms of attacks and defenses, in particular of looking for ways that things could go wrong», Halderman, 2012). It usually takes only one person who sees the exploit («smart cow problem»).
- **Not creating something:** Creativity is worthy of recognition, but even more so is the sacrifice — of ideas, effort, time, and material — when a project is stopped because the dangers are too large. While the rationalization «if I don't develop it, others will» might be comforting, ask yourself: Would you want others to develop it and use it yourself?
- **Ethical (Attention) Design:** Products should promote individual freedom, focus, and agency. That is the opposite of digital crack — apps that make users dependent, capture them with doom-scrolling, small dopamine hits, and the like. See Box 7: Attention Design.
- **Look for unintended outcomes and**

misuse: The aim of the product is usually clearly defined and easy to assess. But what are the outcomes that were not intended? The side effects and far-reaching effects? Imagine the project is finished. What would happen next? What can it be used for? How can it be misused? How will it be understood? Will others see the dangers? Can others understand the dangers? For example, using tablets in school makes it easier to show current information, animations, sounds, etc. But they can also distract students.

- **Dual-use potential:** A specific case of unintended outcomes: The work can plausibly be used for both beneficial and harmful purposes, and the harmful use would be easier, more scalable, harder to detect, or harder to reverse because the work exists. Dual-use potential raises a challenging question: What concrete safeguard reduces likely misuse without destroying the legitimate use? It can require exploration of misuse scenarios, access controls, staged release, external review, documentation limits, monitoring, or, in extreme cases, non-release.
- **Monitor the Effects Long-Term:** Some dangers appear only over time so follow up on your creative projects. You do not know what might happen with them when you create them, but their effects will probably become very obvious with time. If they are negative or if someone uses the work to produce negative effects, take steps to intervene. After all, you are its creator and share responsibility.
- **Emotions:** Negative emotions have their

place, e.g., anger or sadness (cf. «sad film paradox»). However, provide the user with a way out of negative emotions if needed.

- **Ethical, privacy, and security implications are hard or impossible to foresee:** Technology is not only value-neutral; developers are often surprised by how their products are used. Products are often used for other purposes, sometimes more successfully so. Often products rely on other services that create additional issues. For example, an app should be anonymous because it does not create logfiles. However, it runs on a server that does create them, and the IT department has the IP addresses.
- **Digital Data is extremely volatile:** The physical world is hard to surveil; in the digital world, surveillance is part of the infrastructure. Digital interactions leave traces — logfiles, metadata, sometimes clear records. The individual use is hard to predict, and even worse are the interactions. It is often impossible to foresee how different data sources can interact. For example, correlations between different behaviors, such as subscriptions to the adult channel skyrocketing after a religious group took over the hotel for a convention. While correlation is not causation, it can hint at interesting patterns. It also reveals the gap between what people do and what people should do, at least according to different ideologies.
- **Persuasive Products:** Relevant for habit change and all products that aim to change the user.

Box 7: Attention Design

Inspired by Tristan Harris' Time Well Spent work. Use when the work depends on repeated engagement, notifications, feeds, recommendations, streaks, likes, autoplay, or social comparison.

Ask:

- What does the user want before opening the product?
- Does the product help them finish and leave?
- Which features extend use beyond the user's intention?
- Can the user disconnect without penalty, anxiety, or loss?
- Are notifications serving the user's timing or the product's metrics?
- Does the product create competence, or dependence?
- Does it strengthen real relationships, or substitute metrics for them?
- What would we remove if we optimized for user agency rather than engagement?

The point is not to make technology pleasant. The point is to prevent useful tools from becoming attention traps. If the product's business model depends on the user losing track of their own intention, the design has already drifted.

User Agency: If habit change is the aim, the user decides whether to start and how to change. Persuasive technology, gamification, and nudging must be done with

Ethics in Creativity Worksheet

informed consent. Stopping or pausing must be available. «Secret consumption» (not tracking specific behaviors) must be possible.

Broad measures: If user behavior and experience are captured, the measures must reflect the whole phenomenon. For example, reducing fitness to BMI is easy, but deceptive. Never reduce the whole person to one value.

Positive Emotions only: While negative emotions have a long history in behavior change (e.g., electroshocks), they come with unintended effects, including avoiding parts of the situation or learned helplessness if done wrong. Use positive emotions unless the life of the person is in danger (e.g., slapping the hand of a child trying to reach into a power outlet).

Target Audience Feedback

A lot of creative works are developed with feedback from the target audience (e.g., user studies), especially in iterative development. In these cases, treating participants ethically is not only an ethics issue, it also builds trust. Thus, it makes it more likely that you get usable feedback and avoid burning the audience.

In any case, user studies should only be conducted if you are actually willing to use the results and the results themselves can influence the development of the work. They should not be used merely to confirm what you assume — that would waste the time and effort of the participants.

The following aspects can be useful for conducting user studies ethically. Depending on

where you work, stricter limitations by ethics commissions or legal restrictions might apply.

- **Respect:** While this point sounds trivial, lack of respect up to dehumanization can happen easily if a lot of effort was invested into the work and it fails badly with users. Expressions such as PEBCAK (problem exists between chair and keyboard) or DAU («dümmer anzunehmender User», German expression for «most stupid assumable user») are cute and an important coping strategy in tech support, but poisonous in development. It does not solve the problem; worse, it misplaces it on the user side.
- **Informed Consent:** Participants must know in advance what is happening and voluntarily agree to it. To be voluntary, potential participants must have the option to decline without facing negative consequences, especially if they are in dependent positions (e.g., subordinates, students, etc.). Informed consent does not guarantee that users make the best decision — depending on how «best» is defined — but they have the largest interest in making a good decision and are affected the most by the consequences of that decision (Athanasoulis, 2002). In more complex user studies, provide potential participants with written information and ask for their signature if they agree. See Box 8 for a Consent Script.
- **Confounding Factor — Compensation:** On the one hand, compensation is to be expected; participants are spending their time. On the other hand, compensation can skew voluntary participation. For

some groups, e.g., people who are unemployed, homeless, or students, even «low» amounts of money can make refusal too costly.

- **Abort Without Negative Consequences:** Even if people have agreed to participate, they must be able to abort participation at any time — and be informed of that option beforehand. That option must never be questioned. The compensation is paid proportionally to the time of their participation.
- **Removal of Data:** Participants can request removal (deletion) of their data until a specified time after participation (e.g., one week, via an anonymous code that is connected to their data). If they request it, you have to delete their data.
- **Debriefing:** Participants must be informed at the end of their participation about the goals of the study. Negative effects must be reduced, e.g., if participants became frustrated. The camping/relationship rule applies here — leave in a better state than how you encountered them.
- **Anonymity:** Participation is usually anonymous (but see Co-Design). The name of the participant does not matter, but categories can be relevant, e.g., sex, age, expertise, etc. Ensuring anonymity can be surprisingly hard — answers in free text fields, rare combinations of demographic variables (e.g., older male psychology student), etc. can allow identification. Anonymize the data, especially if the answers could have negative consequences for participants (e.g., mistakes at work,

DRAFT VERSION FOR FEEDBACK

Box 8: Consent Script

Purpose: make participation genuinely voluntary and informed.

Use when: testing with users, readers, participants, students, employees, clients, children, vulnerable people, animals, or anyone in a dependent position.

Basic script:

- You are being asked to take part in [activity/study/test].
- The purpose is to learn [purpose].
- You will be asked to [tasks], which should take about [time].
- Possible discomforts or risks are [risks].
- You do not have to participate. You can stop at any time without giving a reason and without negative consequences.
- We will collect [data]. The data will be used for [use], stored [where/how], seen by [who], and deleted after [duration].
- [If relevant:] Audio/video will be recorded only if you agree.
- You can ask questions now or later.
- Do you agree to participate?

Keep it short. If the script needs three pages, the situation probably needs formal review.

risky driving). Confidentiality also applies to who did and did not participate in the study.

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

- **Co-Design:** In some cases, participants are actively asked to improve the design, e.g., in co-design, co-creation, etc. In these cases, participants might expect that their contribution is attached to their name. As this conflicts with the anonymity requirement, ask participants whether they want this.
- **Deception:** It is possible to conceal some information about the study if that information would make the data unusable. For example, an app that is supposed to increase prosocial behavior could not be tested for its effects if participants knew that goal beforehand. They would try to be «a good participant» and show the behavior because of the user study, not because of the app itself. However, they need to be informed about the tasks, time, etc. Deception is not allowed if physical pain or strong emotions are expected. After the relevant part of the study is over, participants must be informed about the deception (debriefing) and told that they can retract their data without negative consequences. If done well, few if any will request that. High retraction numbers indicate a significant ethical issue.
- **Children as Participants:** If the target audience is not of age, both the parents and the children must be informed and voluntarily agree. Neither one is sufficient on its own. Even small children can be informed of the tasks in advance, even if the purpose is beyond their understanding.
- **People with Agency Constraints:** If potential participants have mental disorders,

cognitive impairment, or other issues that reduce agency, their legal guardian has to agree to participation as well.

- **Animals:** While animals are limited in understanding and responses, non-verbal cues should be heeded if you develop products for them.
- **Audio/Video Recordings:** Recordings can provide in-depth data; however, they can also easily identify participants. Treating them confidentially and anonymizing them is usually required here (e.g., transcribing them, including with AI). Explicit written agreement to record the user study is strongly recommended.
- **Surveys:** Surveys are commonly used in analysis (e.g., Is there a need? What are people currently doing?). If the study consists of people filling in surveys, then their participation can be seen as agreement. However, participation must not lead to negative consequences for them, i.e., the data must be anonymized.

Trial Definition

If you want your work to be more ethical, look at the different areas: Ethical Standards, Creative Works, and Target Audience Feedback. Are there areas where your work might be effective and efficient, but perhaps also a bit «fast and loose»? Looking at your last projects and doing an ethical post-mortem might work here.

Choose one project. Use a current project, a recent release, or a recurring work practice.

You also find Boxes with focused issues on this worksheet (see also Table 1):

| Ethical pressure point | Best artifact |
|------------------------------|-----------------------------------|
| Pressure to compromise | Pressure-Tested Red Lines |
| Possible harmful use | Misuse Table |
| Data collection/storage | Data Inventory |
| User agency/control | Stop / Control / Delete Checklist |
| Participant testing | Consent Script |
| You are attached or biased | External Review Request |
| Effects appear after release | Post-Release Monitoring Plan |
| Project may need stopping | Kill Criteria |

Table 1: Ethical Pressure Points

- Box 9: Risk Triage
- Box 10: Pressure-Tested Red Lines
- Box 11: Kill Criteria for a Project
- Box 12: External Review Request
- Box 13: Misuse Table
- Box 14: Data Inventory
- Box 15: «What the User Can Stop / Control / Delete» Checklist
- Box 16: Post-Release Monitoring Plan

Do not complete all artifacts. Choose one artifact that matches the ethical pressure point. The artifact is not the trial. The trial is whether using it changes a decision, design choice, release condition, or working practice.

For low-risk projects, choose one artifact and one concrete change. Do not build an ethics bureaucracy around a small decision.

(see Table Ethical Pressure Points)

Name the concrete weakness. For example,

«The prototype collects more data than needed.», «Users can start but not easily stop.», «Feedback participants are not told how their input will be used.», or «Nobody outside the team has looked for misuse.»

Choose one intervention. For example:

- remove one data field
- add a withdrawal/deletion option
- rewrite consent text
- add an opt-in default
- ask one independent reviewer to attack the design
- create a misuse table
- define one red line that stops release
- add one post-release monitoring check
- change one incentive that rewards cutting corners

Use the Integration Worksheet: Turn the intervention into observable behavior with

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

trigger, start action, behavior, time source, success/abort criteria, trial duration, and evidence. Trials must rely on what actually happens, not feelings or intentions.

Ethics Failure Modes

Ethics is especially vulnerable to becoming reflection, identity, or verbal commitment instead of changed practice. It is not enough to «think ethically» and act badly, so look for concrete changes, e.g., a changed consent form, a removed dark pattern, a data-minimization rule, an external review, a misuse table, or a red-line decision.

Typical failure modes are:

- **Ethics Theater:** more documents, no changed decision.
- **Moral Licensing:** one good feature excuses a harmful pattern.
- **Compliance Substitution:** legal approval replaces ethical judgment.
- **Precaution Paralysis:** no action because not all consequences are knowable.
- **Checklist Tunnel Vision:** only listed harms are considered.
- **Outsourced Conscience:** commission/consultant approval ends thinking.

Example: FocusLock App — Good Intentions, Reduced Agency

Project: A focus app that helps users avoid distraction by blocking selected apps and websites during planned work sessions.

Intended benefit: The user wants fewer interruptions, less doomscrolling, and more uninterrupted creative work.

Ethical pressure point: User agency.

Concrete weakness: The app protects focus by making interruption impossible. Once a session starts, the user cannot pause, stop, or access blocked tools without losing the session streak. The design assumes that the user's future self must be protected from the user's present self. This may help with distraction, but it also turns support into control.

Good intention: Help the user defend attention.

Ethical risk: The app may reduce autonomy, punish legitimate changes of plan, create dependency, and make the user feel managed rather than assisted.

Chosen artifact: «What the User Can Stop / Control / Delete» Checklist.

Intervention: Add a pause/override system that preserves agency without making avoidance frictionless.

Artifact Result: Stop / Control / Delete Check

See Table 2.

Main finding: The app currently treats failure to follow the plan as a moral problem. It should treat it as information.

Trial Definition

Reason for Change: The app is designed to protect focus, but the current blocking logic may overcorrect. Users can start a focus session easily, but they cannot pause, stop, or override without penalty. This may improve adherence while reducing agency.

Current Pattern: During a focus session, blocked apps and websites cannot be

accessed. Stopping early breaks the streak. Override attempts are logged. Users can technically quit the app, but the interface frames this as failure. If nothing changes, the app will continue to support focus by making the user less able to adapt.

Goal Behaviors: The app should preserve focus while keeping the user in control. Users should be able to pause, stop, or override a session for legitimate reasons without being punished or moralized. The change must not make blocking meaningless.

Change Behaviors

Trigger: Before implementing any new blocking or commitment feature.

Start Action: Open the Stop / Control /

Delete Checklist.

Behavior: Check whether the feature allows pause, stop, override, data deletion, and non-punitive exit. Any «No» requires either redesign or explicit justification.

Time Source: Replaces ten minutes of feature-planning time during design review.

Intervention to Test

For the next version, add:

- pause mode without streak loss
- emergency override without shame language
- optional reason tagging after override
- weekly review of override patterns
- clearer data deletion/export controls

| Question | Current state | Change needed |
|--|---|--|
| Can the user stop using it without penalty? | No. Stopping breaks streak. | Add stop option without shame language. |
| Can the user pause it temporarily? | No. | Add pause mode with reason tagging. |
| Can the user leave without losing unrelated data/access? | Mostly yes. | Make export clearer. |
| Can the user delete their data? | Hidden in settings. | Move to account/data section. |
| Can the user turn off notifications/prompts/recommendations? | Partly. | Add global notification control. |
| Can the user reverse mistaken actions? | No, blocked session cannot be changed. | Add emergency override. |
| Can the user use the core function without unnecessary surveillance? | Partly. App tracks all override attempts. | Track only aggregate override count unless user opts in. |

Table 2: Artifact Result: Stop / Control / Delete Check

DRAFT VERSION FOR FEEDBACK

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Success Criteria

The redesign counts as successful if:

- users can pause or stop a session without shame language or streak punishment
- blocked access still requires a deliberate action, not one accidental tap
- the app stores only minimal override data by default
- at least one design decision is changed because of the checklist

Abort Criteria

Abort or redesign the intervention if:

- pause/override makes the blocker functionally useless
- users are still punished for stopping
- the team adds new tracking to compensate for lost control
- the checklist becomes documentation only and changes no design decision

Ambiguity: An override does not automatically count as failure. It counts as information. Repeated overrides may mean the session design is too rigid, the blocked category is too broad, or the user's work context changed.

Trial Duration: Two design cycles or four weeks, whichever comes first.

Evidence to Log

- checklist completed for each blocking-related feature
- design changes caused by checklist
- number of features released unchanged / redesigned / removed

- user complaints about feeling trapped, punished, or overcontrolled
- user complaints that blocking became too easy to bypass

Decision

- **Keep** if the checklist changes design decisions while the app still supports focus.
- **Modify** if agency improves but the blocker becomes too weak.
- **Abort** if the checklist becomes ethics theater or the design still depends on coercion.

Reflection Prompt: Did the app become better at supporting agency, or did it merely hide control behind softer language? Did the user gain meaningful control, or only more settings?

Next Step: If kept, make the Stop / Control / Delete Checklist mandatory for every feature that restricts behavior, access, data, or communication.

Hand-Off

Do not try to become «an ethical person» in this worksheet. Pick one project, one recurring decision, or one artifact where harm, coercion, deception, dependency, waste, or misuse might enter. Change that. Then test whether the change survives real work pressure.

What will you do differently because this artifact exists?

More Information

- Athanassoulis, N. (2002). The Role of Consent in Sado-Masochistic Practices. *Res Publica*, 8, 141–155.
- Deci, E. L., & Ryan, R. M. (2000). The “What” and “Why” of Goal Pursuits: Human Needs and the Self-Determination of Behavior. *Psychological Inquiry*, 11(4), 227–268.
- Duke, A. (2022). *Quit*. Random House Books.
- Holiday, R., & Hanselman, S. (2016). *The Daily Stoic. 366 Meditations on Wisdom, Perseverance, and the Art of Living*. Portfolio.
- Schneier, B. (2006, January 12). Anonymity and Accountability. *Schneier on Security*. https://www.schneier.com/blog/archives/2006/01/anonymity_and_a.html

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 9: Risk Triage

Classifying the ethical risk can be a useful step: low / medium / high / stop-and-escalate. Classification depends on reversibility, number of affected people, dependency, data sensitivity, vulnerability of users, coercion, safety-critical effects, public release, and inability to withdraw.

Escalation Conditions

Stop and consult someone when the work affects:

- vulnerable participants,
- sensitive data,
- deception,
- surveillance,
- coercion,
- safety-critical use,
- children,
- medical/psychological effects,
- irreversible release, or
- dual-use potential.

Box 10: Pressure-Tested Red Lines

Purpose: define boundaries before pressure appears.

Use when: money, deadlines, authority, ambition, publication pressure, audience growth, reputation, attachment, or survival pressure might tempt compromise.

Format:

- We do not ...
- Especially when ...
- The likely rationalization will be ...
- If this happens, the decision moves to ...

Ethics fails when it is written for calm conditions but used under pressure. Red lines need enforcement design. Otherwise they are only statements of principle.

| Red line | Especially when | Likely rationalization | Required safeguard |
|--|---|--|--|
| We do not hide relevant limitations. | A release, grant, sale, publication, or loved person depends on success. | «This is temporary.» / «The good outweighs it.» / «They would not understand.» | External reviewer must approve disclosure before release. |
| We do not use participants merely as means. | The data would otherwise be unusable. | «It is only a small deception.» / «Everyone does this.» | Predefine concealment limits, debriefing, and data-removal conditions. |
| We do not punish critics. | The criticism threatens reputation, funding, authority, or someone close to us. | «They are harming the project.» / «They are bad-faith.» | Criticism is logged and answered before any retaliatory action. |

A red line is only real if violating it changes who decides, what must be disclosed, or whether the project continues.

Box 11: Kill Criteria for a Project

Purpose: decide in advance when stopping is the ethical action.

Use when: the project is seductive, expensive, prestigious, personally meaningful, externally pressured, or has sunk costs.

Format: We stop or pause this project if ...

Possible kill criteria:

- The project requires deception we cannot justify or repair.
- The project depends on coercion, dependency, or user helplessness.
- The intended benefit can be achieved by a less harmful method.
- A serious misuse is likely and cannot be mitigated.
- Users cannot meaningfully understand, refuse, stop, or delete.
- Harm would be irreversible or difficult to detect.
- Data exposure would endanger users.
- The project works only by hiding relevant information.
- The project's incentives reward harm, addiction, surveillance, fraud, or manipulation.
- External reviewers independently identify the same serious risk.
- We keep needing «good intentions» to defend the project.

Killing a project is not failure if the project should not exist in its current form.

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 13: Misuse Table

Purpose: make likely harmful uses visible before release.

Use when: the work increases power, scale, persuasion, automation, access, anonymity, surveillance, dependency, or damage potential.

Format:

| Feature / capability | Intended use | Plausible misuse | Who could misuse it? | What harm becomes easier? | Safeguard | Residual risk | Decision |
|----------------------|--------------|------------------|----------------------|---------------------------|-----------|---------------|----------|
| | | | | | | | |
| | | | | | | | |

Prompts:

- Who benefits if this is abused?
- Who is harmed if this works too well?
- Does this lower cost, increase scale, improve precision, or reduce accountability for harmful action?
- Could the same feature be used by a careless user, hostile actor, desperate organization, activist, criminal, state actor, or employer?
- Can misuse be reduced by access limits, friction, logging, staged release, documentation limits, monitoring, or redesign?

Decision options: release unchanged, release with safeguard, release only to limited users, delay release, remove feature, or abandon project.

Box 14: Data Inventory

Purpose: prevent «just in case» data collection.

Use when: the project collects, stores, combines, infers, shares, profiles, predicts, or logs user behavior.

Rule: No data without a job. No retention without a reason. No sharing without explicit justification.

Format:

| Data item | Why needed? | Required or optional? | Sensitivity | Who sees it? | Stored where? | Kept how long? | Shared with? | Worst plausible harm | Decision |
|-----------|-------------|-----------------------|-------------|--------------|---------------|----------------|--------------|----------------------|----------|
| | | | | | | | | | |
| | | | | | | | | | |

Good decision options: do not collect, collect only locally, make optional, aggregate, anonymize, shorten retention, separate from identity, delete after use, explain more clearly, or require explicit opt-in.

Box 12: External Review Request

Purpose: get someone outside the project to look for harm.

Use when: you are attached to the project, the stakes are high, the audience is vulnerable, the work is persuasive, data-heavy, dual-use, public, or hard to reverse.

Do not ask someone whose incentives depend on approving the project. That includes friends or family members who might want to preserve the relationship.

Request template:

- I am developing [project].
- Intended use: [intended use].
- Target users/audience: [who].
- Main benefits: [benefits].
- Main concern: [concern].

Please review it with a hostile/careless/misuse mindset.

I am especially asking you to look for:

- ways this could harm users
- ways this could be misused
- hidden coercion or dependency
- data/privacy/security issues
- places where my incentives distort my judgment
- reasons not to release

I need your blunt assessment, not reassurance. There will be no negative consequence for telling me to stop, redesign, or kill the project.

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 15: «What the User Can Stop / Control / Delete» Checklist

Purpose: protect user agency.

Use when: the work affects behavior, attention, data, access, identity, communication, visibility, reputation, or dependency.

Core question: Does the user remain an agent, or are they being managed?

Checklist:

| Question | Yes / No | Change needed |
|--|----------|---------------|
| Can the user stop using it without penalty? | | |
| Can the user pause it temporarily? | | |
| Can the user leave without losing unrelated work/data/access? | | |
| Can the user delete their data? | | |
| Can the user export what belongs to them? | | |
| Can the user turn off notifications/prompts/recommendations? | | |
| Can the user understand what is being collected? | | |
| Can the user opt in rather than merely opt out? | | |
| Can the user reverse mistaken actions? | | |
| Can the user appeal moderation/automated decisions? | | |
| Can the user use the core function without unnecessary surveillance? | | |

Box 16: Post-Release Monitoring Plan

Purpose: responsibility does not end at release.

Use when: effects may appear only after use, scale, interpretation, misuse, dependency, or combination with other systems.

Monitoring is not reputation management. It is checking whether reality contradicts the creator's assumptions.

Format:

| What to monitor | Signal | Source | Check frequency | Threshold | Action |
|-----------------|--------|--------|-----------------|-----------|--------|
| | | | | | |
| | | | | | |
| | | | | | |

Examples:

- user complaints about manipulation, confusion, dependency, harm
- unexpected user groups adopting the work
- data breaches or suspicious access patterns
- misuse cases appearing in the wild
- user drop-off after coercive friction
- negative effects on attention, autonomy, competence, relationships
- public misunderstanding of the work's limits
- support requests that reveal hidden harm

Decision thresholds:

- if X happens once → investigate
- if X happens three times → redesign
- if X affects vulnerable users → pause release
- if X causes serious harm → disable feature
- if misuse becomes common → restrict access or withdraw

Ethics in Creativity Worksheet

DRAFT VERSION FOR FEEDBACK

Box 17: Good Intentions Quotations

«The cultivation of good intentions is an extremely undemanding intellectual activity. Designing plans to realize noble goals is a different matter. This requires intelligence. The appreciation of good intentions alone is by no means appropriate, on the contrary!»

In my opinion, it is an open question whether <good intentions + stupidity> or <bad intentions + intelligence> have brought more mischief into the world. Because people with good intentions usually have few inhibitions about tackling the realization of their goals. In this way, inability, which would otherwise remain hidden, becomes dangerous, and in the end there is the astonished and desperate exclamation: <We didn't want this!>»

Dörner (2015, translated)

«Of all tyrannies, a tyranny exercised for the good of its victims may be the most oppressive. It may be better to live under robber barons than under omnipotent moral busybodies. The robber baron's cruelty may sometimes sleep, his cupidity may at some point be satiated; but those who torment us for our own good will torment us without end, for they do so with the approval of their own conscience.»

C. S. Lewis

«No person is more capable of evil than the ones convince they can only do good.»

Unknown

«If only it were all so simple! If only there were evil people somewhere insidiously committing evil deeds, and it were necessary only to separate them from the rest of us and destroy them. But the line dividing good and evil cuts through the heart of every human being. And who is willing to destroy a piece of his own heart?»

During the life of any heart this line keeps changing place; sometimes it is squeezed one way by exuberant evil and sometimes it shifts to allow enough space for good to flourish. One and the same human being is, at various ages, under various circumstances, a totally different human being. At times he is close to being a devil, at times to sainthood. But his name doesn't change, and to that name we ascribe the whole lot, good and evil.»

«The Gulag Archipelago» by Aleksandr Solzhenitsyn

«Sure as I know anything I know this, they will try again. Maybe on another world, maybe on this very ground swept clean. A year from now, ten, they'll swing back to the belief that they can make people better.»

Capt. Malcolm Reynolds in «Serenity»

«My father used to say, that the road to hell is paved with good intentions. I laid the first stone right there. I'd committed myself. I'd pay any price, go to any length, because my cause was righteous ... my intentions were good. In the beginning, that seemed like enough.»

Star Trek DS9: «In The Pale Moonlight»

«Good intentions don't matter. Only thing that matters is behavior, not intentions. Almost nobody has evil intentions. Very few people wake up in the morning and say, <Oh, another opportunity to do evil.> Stalin didn't think he was evil. Hitler didn't think he was evil. Mao didn't think he was evil. Okay, just for the record. They all thought they were good. Communism, the greatest mass murdering machine in history, killed a hundred million civilians in the 20th century alone, was based on good intentions. Equality, fraternity, anti-imperialism and anti-colonialism and whatever else. Good intentions don't mean anything. You should judge yourself and everyone else by their behavior.»

Dennis Prager

«... in that book, there is a scene where there are people who oppose this alien thing. And they say, we mustn't use this. We mustn't touch this. We need to get these aliens out of here. And another guy says to him, why do you fight these aliens? Like, they want good. They're trying to help us. They're trying to improve things. And he says, I don't fight them because I think they're going to do evil or they want to do evil. I fight them because they're going to do good as they understand it. And that's why power is dangerous, because everyone can say, I don't want power to do bad things. I want power to do good things. Now give it to me.»

Konstantin Kisin about the aliens in «Stalker» on Joe Rogan Experience 2297

Bonitas non est pessimis esse meliorem.
[It is not goodness to be better than the worst.]

Seneca

«The surest way to work up a crusade in favor of some good cause is to promise people they will have a chance of maltreating someone. To be able to destroy with good conscience, to be able to behave badly and call your bad behavior 'righteous indignation' — this is the height of psychological luxury, the most delicious of moral treats.»

Aldous Huxley

««The lesser of two evils», or «the greater good». Get a good man to utter either of those phrases, and there is no one more eager to begin perpetrating evil.»

*«An Echo of Things to Come»
by James Islington*