



FOR HUMANITY

THE RISE OF THE ETHICS COMMITTEE:

APR 2021 // PREPARED BY RYAN CARRIER

ACKNOWLEDGMENTS

The author would like to thank Jacquie Hughes, and Mark Potkewitz, for their invaluable feedback on earlier drafts of this report. Special thanks to Tristi Tanaka and the #ethicscurriculum team and the #children's code team at ForHumanity fo helping to shape and develop these ideas.

PRINT AND ELECTRONIC DISTRIBUTION RIGHTS



© 2021 by ForHumanity. This work is licensed under a Attribution-NonCommercial-NoDerivatives 4.0 International license.

To view a copy of this license, visit: <https://creativecommons.org/licenses/by-nc-nd/4.0/>



Ryan Carrier, CFA

Ryan is Founder, Executive Director, and Chairman of the Board of Directors of ForHumanity.
Ryan@Forhumanity.center | <https://forhumanity.center/>



Is
human agency
slipping away?

Introduction

A NECESSARY RESPONSE TO THE SUSTAINABILITY OF PROFIT THAT DOESN'T COME AT THE EXPENSE OF HUMAN AGENCY.

It is difficult to pinpoint when the trend started. It may be linked to global concern about climate change — a problem too existential to handle by any single person, company, or even nation-state. Or maybe, well-publicized incidents of racial injustice may have accelerated a sense of a power and accountability imbalance. Or could it be related to a growing unease about the impact of pervasive technology on our individual and collective well-being played its part. Whatever the root cause, source, or spark, there is no doubting the gathering demand for accountability, oversight, and governance of all institutions and entities who have any kind of domain over our lives. The goal is worthy — to protect humans and human agency. However, the “soft” terms being applied, such as well-being, ethics and good social order, present new challenges for adherence, compliance, and enforcement.

In an attempt to control and contain rapid and encroaching technological development, soft laws are being developed at pace, to aid hard laws in attempting to protect and sustain human agency. But we should not expect laws to do all the heavy lifting: company Boardrooms and C-suite officers - faced with critique and criticism about the impact of their products and services on an unsuspecting public - must do their share. At the time of writing, some companies have dug in their heels while others are beginning to search for viable long-term solutions for the impact of algorithmic technology being ubiquitously advanced often at the expense of humans.

Companies are deploying systems throughout every aspect of the company, affecting humans at every turn. Defending humans and human agency in the midst of substantial societal upheaval, and technological change requires a unique set of skills and those skills cannot be applied post hoc, but must be applied throughout an algorithm's lifecycle. A hallmark of recent technological change is the increased ability of algorithms to nudge, guide, choose, interpret and decide "for us", and this begs the question: which elements 'of us' are being encoded? Many of us recognize when our shopping preferences have been encoded into recommendation lists, but what about more subtle examples of our encoded morality, priorities, well-being? And related to all of those crucial questions, where is 'our agency' in the design, development and application of these systems?

As company designers and developers build more and more algorithmic systems designed to replicate/replace portions of human decision-making or physical activity, they tread into areas of well-being and moral frameworks for which they are ill-equipped. Worse still, they sometimes lack the awareness of the ethical impact of their work that verges on recklessness. Design teams solving problems, but also assessing the associated risks of their work is, at best, a conflict-of-interest and certain to be filled with myriad cognitive biases which prevent a fair or reasonable evaluation of risk. Failure to manage the risks from diminished human agency and failed application of moral frameworks can have devastating consequences for people.

Examples of Soft Law

Recently, government and regulatory response have created a few protections to reclaim lost agency — a legal one, the [General Data Protection Regulation](#) (GDPR), is filled with soft law descriptions like the one found in article 22.1 (amongst others articles, the emphasis is added):

The data subject shall have the right not to be subject to a decision based solely on automated processing, including profiling, which produces legal effects concerning him or her or similarly significantly affects him or her.

The Rise of the Ethics Committee



HARD VERSUS SOFT LAW

. Also, in the UK, the [Age-Appropriate Design Code](#) (hereafter “Children’s Code”), in response to online harms against children, became law. The Children’s Code chooses to reinforce the United Nations Convention on the Rights of the Child (UNCRC) for those subject to the law. It borrows the definition of Best Interests of the Child, which is overflowing in soft law language:

The UNCRC incorporates provisions aimed at supporting the child’s needs for safety, health, wellbeing, family relationships, physical, psychological and emotional development, identity, freedom of expression, privacy, and agency to form their own views and have them heard. Put simply, the best interests of the child are whatever is best for that individual child.

In 2019, after repeated criticisms of a “profit-only” raison d’être, the [Business Roundtable](#) addressed the role of corporations in society with this amendment to the statement of purpose of a corporation:

the release of a new Statement on the Purpose of a Corporation signed by 181 CEOs who commit to lead their companies for the benefit of all stakeholders – customers, employees, suppliers, communities, and shareholders.

Risks from Soft Law and Artificial Intelligence/Algorithm Risk

Collectively, these changes amount to strong support for the well-being of humans and human agency. However, figuring out how to facilitate and implement these changes has eluded many corporate leaders to-date, leaving the world with the impression of “hot air” rather than intentional change. This paper aims to provide a specific action plan for CEOs who either must comply (when the soft law is enacted) or desire to comply (when they choose to abide by statements like those of the Business Roundtable). We hope that a roadmap and process of oversight will enable corporations to “meet the promise” from the Business Roundtable.

We recognize the theme of this paper flies in the face of a recent news items from Google; ethicists [Timnit Gebru](#) and [Meg Mitchell](#) were both let go which dealt a significant blow to the tech company's ability to meet the challenges of tackling the risks and soft law implications across their vast product suite. Instead, Google continues to handle these risks, like most other corporations with the mistaken belief that their business managers, legal team or designers and developers, without training, can navigate questions of ethical choice and compliance with soft law. We expect that this belief will prove costly — far greater than the cost of the wise development and deployment of an empowered, integrated and well-considered ethics committee.

We believe that Google's recent actions, including the failed introduction of an [ethics panel](#), will be looked back at as a watershed moment and highlighted as a key reason for the **Rise of the Ethics Committee**. “The Rise” reflects the present state of the ethics committees and our opinion that they currently remain mostly ineffectual. Today's ethics committees are the equivalent of “window dressing” — not because of the people involved but rather a failed corporate commitment to adherence and accountability with respect to their own moral frameworks or sometimes worse: disregard for human agency. Called upon to draw up codes of conduct, codes of ethics, and recently, codes of data ethics, these teams are cobbled together, usually with a strong concentration in legal compliance. Some of the employees named in these committees may be well-intentioned and sincere, but far from being ideal choices, they are far too often the least worst choices. For decades, whatever these teams decide has been the definition of “ethical” for corporations. These unintegrated teams, many of whom have little training in Ethics, Applied Ethics, or Philosophy, are treated as afterthoughts when compared with the primary objective of maximizing shareholder value. Suffice it to say, most ethics committees would not have to improve much to represent a dramatic “rise” in their materiality.

Soft Law Explained

While we use the term “law” to refer to enacted statutes with clear terms and definitions, we consider soft law to include things like regulatory guidance, commentaries from persuasive authorities like regulators or government agencies, or guidelines published by official governmental sources. Below is the most recent definition we could find, it is not current:

Soft law refers to rules that are neither strictly binding in nature nor completely lacking legal significance. In the context of international law, soft law refers to guidelines, policy declarations or codes of conduct which set standards of conduct. However, they are not directly enforceable. Hard law refers to binding laws.

GDPR and the Children’s Code in the EU and UK respectively are binding laws and are directly enforceable. Therefore if we use the former definition and bring it current, we can manufacture a more current definition for “soft” law.

Soft law refers to rules that are neither strictly binding in nature nor completely lacking legal significance. In the context of international law, soft law refers to guidelines, policy declarations or codes of conduct which set standards of conduct. However, it is not well known or adjudicated on how these guidelines, policy declarations and codes of conduct will be enforced by the judicial system.

Our estimation is that the legal application of these “soft” law criteria will likely fall into the same category as “reasonable, fiduciary or industry-standard”. Companies will need to have evidenced-based documentation to support claims that they are in compliance with the law. In addition, expertise will be required to satisfy compliance standards as qualified and accountable. Unfortunately when it comes to the “best interests of the child” or determining the impact of profiling on a data subject - a GDPR requirement - new skills are needed to meet these challenges and robustly manage the risks arising from “soft law”.

The Rise of the Ethics Committee

BIG DATA IS
WATCHING YOU

Risks associated with algorithms and soft law are present, but corporate awareness and management of these risks appear absent. Current risk management solutions rely on strategies such as “hoping to avoid scrutiny” and “embedding the estimated costs of legal liabilities” into the cost structure of the product or service. Neither of these solutions offers long-term viability, especially when companies facing a rising clamor for increased human agency, compliance, and regulation. We suspect that companies will be forced to adopt risk management techniques rather than burying their heads in the sand. Ambivalence or neglect for people’s rights and freedoms is not a sustainable business model and the result will either be consumer backlash or heavy regulatory oversight.

The Rise Starts Now

So how can we explain why the “rise” will begin now? Risk is the answer. At their core, corporations are profit-making ventures, and any risk to profit becomes a priority. As awareness of these risks increases, business managers will seek the most qualified experts for managing that risk — in this case, well-trained ethics officers. Companies are learning firsthand, from new laws, from multi-stakeholder feedback, and from failure to mitigate significant and monstrous externalities tied to their products that these risks are real and significant. Examples include: 1) [excessive influence on democratic institutions](#); 2) [dangerous dabblings in constitutional and fundamental human rights](#); or, 3) more insidiously the embedded bias in our data and artificial intelligence models as highlighted by director Shalini Kantayya and leading authority Joy Buolamwini in [Coded Bias](#). The risks have become mission-critical, so the sustainability of profits and shareholder value are at risk. Action will follow.

Governments and regulators have responded first. Now, company boards, CEOs, and risk managers are beginning to seek remedies for soft law and algorithm risk. ForHumanity, in one example, identified a mechanism to establish an Infrastructure of Trust for AI and Autonomous systems in the form of a compliance-by-design solution: [Independent Audit of AI Systems](#). The process has established audit rules and adaptations of existing law that embed governance, accountability and oversight across ethics, bias, privacy, trust and cybersecurity for AI and Autonomous Systems. Compliance with government approved audit rules/criteria necessitate explainability, transparency and disclosure combined with human agency in risk assessment and feedback loops tied to design and development.

Ethical Choice - the responsibility of Trained Ethics Officers and the Ethics Committee

ForHumanity designed Independent Audit of AI Systems (IAAIS) to create and enforce accountability, governance, transparency and oversight throughout the life-cycle of AI and autonomous systems from design to decommissioning. A key element of this accountability and transparency is the systematic assignment of instances of Ethical Choice for examination, resolution and disclosure by trained ethics officers wherever those choices occur in the system life-cycle. Unlike most current, ungoverned processes, IAAIS removes the temptation for designers and developers to make decisions that they are not trained for and that are filled with their own conflicts-of-interest, such as failure to ignore sunk-costs and confirmation biases. Turning instances of ethical choice over to trained ethics officers will reduce the risk of conflicted, concentrated and myopic choices that often result in negative social impact and reductions in human agency.

The audit criteria assigns instances of ethical choice to an ethics officer no matter where they are found in the design, development, input and risk assessment phases of the AI and Autonomous lifecycle. Again, normalizing the lexicon, we define **Ethical Choice** as:

Awareness of a set of options to be made In the context of automated and intelligent systems, using a set of principles and rules concerning moral obligations and regards for the rights of humans and for nature, which may be specified by a given profession or group. The result, outcome or judgment is made using a shared moral framework. or set of moral principles based upon the entity's Code of Ethics.

Examples include Min/Max benchmark (discussed in a forthcoming paper) setting and explanations/disclosures for dataset bias remediation or working with designers and developers to establish Key Performance Indicators to assess the stability of an algorithm in accordance with the initial scope, nature, context and purpose. Our current working list of embedded instances of ethical choice in the life-cycle of design and development for an algorithm include:

- 1) Proportionality Assessments
- 2) Necessity Assessments
- 3) Risk Analysis
- 4) Diverse Inputs and Multi Stakeholder feedback
- 5) Reliability tests
- 6) Validity tests
- 7) Explainability

Instances of Ethical Choice occur frequently in the algorithm life-cycle and corporate risk management will benefit from the expertise of trained algorithm ethics officers. This list comes from ForHumanity's GDPR Certification Scheme criteria and in no way reflects the comprehensive instances of ethical choice embedded in the vast array of systems being deployed across the economy. Therefore, it is reasonable and expected that this list/skillset will grow in breadth and depth as IAAIS expands to cover all AIs and Autonomous Systems that impact humans.

The Rise of the Ethics Committee

The aforementioned list is further augmented by soft law examples. Ethics officers need expertise and training to fairly and responsibly consider and evaluate corporate responsibility, priorities and compliance reacting to the expansion of soft law such as:

- 1) The health and well-being of the Child
- 2) Physical, psychological and emotional development.
- 3) Right to develop their own views and identity

These are but a few soft law examples related directly to the Children's Code, however they are representative of the latest evolution in human agency — the further identification and specification of rights and freedoms drafted into law by legislatures around the world.

These examples pinpoint the expertise needed to manage the risks sourced and identified in the audit criteria. Further examples, beyond the scope of an algorithm ethics officer but still in the domain of the ethics committee include areas of societal or existential risk, like ESG (Environmental, social and governance) responsibilities.

Facilitating Genuine Change - a roadmap for Accountability and Transparency

Somewhere between the inevitable consumer backlash and heavy regulatory oversight lies the middle ground of robust protections for humans and a system of compliance, accountability and oversight that will satisfy regulators. We believe that IAAIS, a replica of the system of independent, third-party governance, accountability and oversight which has served the financial accounting and reporting world successfully for 50 years, will supply the needed infrastructure of accountability and transparency to address these risks.

Adoption of certification schemes and audit rules/criteria by regulators and governments creates the opportunity for proactive compliance with laws and best practices to mitigate algorithm risk. Thus we can facilitate genuine change by systematizing accountability. To accomplish any human-led change, accountability is required.

Accountability is the deciding factor between those who talk, wish and hope for change and changemakers who take action and see the act through to completion

It is an unfortunate tenet of human existence that when we are watched or checked-in upon, we are more likely to get the work done or comply with precise requirements. But clearly accountability is a concept that Silicon Valley seems loathe to accept. Like proverbial foxes guarding hen houses, tech companies hire their own ethics panels and require them to apply self-declared standards. To much fanfare and little surprise, these companies find and declare themselves ethical to the highest standards they have conjured.

Accountability is effective and meaningful when it is independent, such as abiding by a set of third-party rules rather than making your own and inviting third-party audits to assure compliance rather than grading your own homework assignments. When companies demonstrate humility to be compliant with representative societal values, consensus-driven best practices, and application of the law, then will they meet this higher standard of accountability. The ethics committee is the agent of oversight, and management of the moral frameworks found in their own code of ethics. But instead of empty words on a page, now, when faced with third-party independent audits, companies will have to submit to compliance audits for societal assurance, and we all will benefit; even the company being audited as their operations will have greater sustainability.

Accountability is further enhanced when the process requires and fosters greater transparency. Now, not only is someone watching, and checking, but the whole world is able to see the work, to review and critique crucial elements of disclosure and compliance. Transparency and disclosure create a feedback loop. Once information is disclosed, critics can provide constructive critique into the process and facilitate future improvements. Accountability combined with transparency creates a virtuous circle of improvement and development.

A fair examination of the accumulation of risks from soft law and instances of ethical choice in our AIs and Autonomous Systems combined with the aspirational appetite of corporations to move beyond prioritizing shareholders at the expense of the humans reaches one conclusion - empowering the ethics committee. Companies will make this choice, not for public relations or ethics washing, but for genuine application of ethical choice, adherence to a moral framework publicly declared in the code of ethics, and integrated into corporate risk management because it will result in a greater likelihood of sustainable profit. Humans benefit with corporate entities that are more attuned to human agency.



The Rise of the Ethics Committee

At some point for sustainability - companies must figure out that they exist because people/humans collectively allow them to exist and choose to use their services. They have rights because people grant them limited rights in our socioeconomic system. These rights are not inherent and unalienable. If corporations continue to negatively influence, alter, degrade or otherwise subvert human agency, those granted rights will be rescinded. The risk of being unsustainable is manifest now.

Sustainability happens when companies accept and embrace The Rise of the Ethics Committee. The alternative, for these entities, appears to be heavy regulatory oversight, and complete transparency at the expense of intellectual property and trade secrets or, worse, antitrust actions which are founded in the explicit recovery of human agency. The right choice seems so obvious — choose human agency, choose humility, choose to empower people to think beyond profit and consider your own declarations in the [Business Roundtable](#) to lead companies for the benefit of all stakeholders — customers, employees, suppliers, communities, and shareholders. Elevating and integrating the Ethics Committee is the key to keeping that promise.

ForHumanity

ForHumanity is a 501(c)(3) tax-exempt public charity formed to examine and analyze the downside risks associated with the ubiquitous advance of AI and automation. To this end, we engage in risk control and mitigation and deploy the lens and filter of Ethics, Bias, Privacy, Trust, and Cybersecurity to ensure the optimal outcome...ForHumanity.

ForHumanity is an interdisciplinary group of dedicated expert volunteers, with over 300 contributors and 32 Fellows, its collective expertise spans the AI field, ranging from ethics to algorithmic risk and to security. Our team is drawn from the academic, legal, policy, corporate, and public sectors of over 40 countries around the world. Our mission is to help create an 'infrastructure of trust' for all autonomous systems that directly impact humans.

ForHumanity drafts comprehensive, pragmatic and implementable audit rules and standards for autonomous systems in every corner of the economy. Our experts collaborate with industry practitioners to ensure these audits achieve our mission of mitigating AI risk to humans. This system of audit rules and standards - adapted to local jurisdictional laws and regulations - is called Independent Audit of AI Systems (IAAIS).