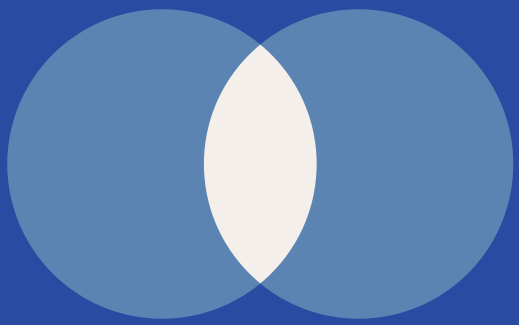
Version 2.0

Corporation for Public Interest Technology

STATE-OF-THE-FIELD LANDSCAPE SURVEY: TECHNOLOGY ETHICS AND PIT PRACTITIONERS



2024

www.cpitpbc.com

The coming years will likely prove decisive for determining how technology can be made more accountable to public interest and societal benefit by expanding public interest technology more fully into the commercial space.

Now is arguably the most urgent time for decisive action on the part of stakeholders committed to public interest technology. The opportunities offered by advancements in technology are tremendous. As this study indicates, the potential benefits of technology innovation are rivaled by the actual and potential harms associated with these same technologies.

These developments have been met, fortunately, by an unprecedented scale of awareness and strategy devoted to foregrounding public wellbeing and making technology accountable to the best interests of society.

SYLVESTER JOHNSON, CO FOUNDER AND CEO

> **Our society is now challenged to create new methods and greater collaboration for democracy, social justice, and sustainability to guide the technological future of humanity.**

I. EXECUTIVE SUMMARY

Public concern about the risks posed by technology—especially, if not exclusively, **artificial intelligence (AI)**—is on the rise. Every day, it seems, a new **op-ed** or **industry report** is published on the need to prioritize "ethics" in the tech sector. This sense of urgency is rooted in the array of ethical challenges that accompany our ever-evolving technologies; challenges magnified by the rapid advancement and widespread integration of AI systems into societal infrastructures. These include issues like data privacy and security, **algorithmic bias**, and digital surveillance, and also extend to the broader impact of technologies on labor, democracy, fundamental **human rights,** and even the environment. There is growing consensus around the social, political, and economic import of incorporating ethical principles and practices into the design, development, commercialization, and use of technology.

What, precisely, **"technology ethics"** means remains unclear, however.

Technologists, journalists, pundits and politicians speak of "aligned," "responsible," and "human-centered" AI. The White House issued an executive order **(AI EO)** in October 2023 calling for **"safe, secure, and trustworthy"** AI. **The European Union's AI Act** (**EU AI Act**) aims to protect fundamental human rights. Public interest and **social justice** technologists not only seek to expand our sense of the harms perpetuated by AI systems—illuminating, for instance, how algorithms can reproduce and reinforce inequalities at unprecedented scale—but also aspire to develop new tools to broaden impacts and produce just outcomes.

While these terms are not mutually exclusive, neither are they equivalent. Each comes with its own assumptions about the world as it is and ought to be. Depending on the context, "tech ethics" might refer to the equivalent of a vocational ethics program one might find in a business school; industry guidelines that protect against reputational and financial damage; academic research aimed at understanding the human costs, environmental impacts, and philosophical implications of new technologies; activist efforts to illuminate discriminatory and oppressive systems and hold tech corporations and state actors accountable; and everything in between.

The **elite capture** of "ethics" discourse has led some to question whether we should speak of "ethics" at all. Without enforceable laws to hold technology companies accountable to the public interest, there are legitimate concerns over **ethics washing** and the corporate capture of the regulatory process. This has led some to suggest we adopt alternate terms like **"power," "human rights,**" or, indeed, "**social justice**." Stakeholders will need to consider the difference between more generic conceptions of ethics and the specific inputs, outcomes, and approaches demanded by a term like **social justice**, which seeks to create a more just and equitable society.

An important distinction emerging in the landscape is the difference between those seeking to produce "responsible tech" products and those who explicitly emphasize the public interest and **social justice** in their corporate mission.

These debates notwithstanding, broadly conceived, tech ethics is defined by two questions:

- How can we prevent, mitigate, and address the harmful impacts of technology?
- How can we use technology to ensure positive outcomes for humanity and the world?

How do we ensure tech does no harm, and how might we use tech to do good? If the diverse stakeholders working in the tech ethics space are to help bring about a better world, they will need to do just that: design, develop, and deploy technologies that enact practices of justice, and enforce limits on the power of technologies and the entities that develop and market them. They will also need to create and sustain a **social justice** tech ecosystem, given the scale and complexity of the challenges presented by LLMs, algorithmic systems, and other technologies.

In an ideal world, tech companies would approach the full lifecycle of their products in ways that ensure they help solve real problems of **sociotechnical** importance, and, at the very least, do not harm or exacerbate harm. Beyond good intentions, however, there will be additional pressures at work in the coming years, not least of which will be compliance with regulatory standards. In addition, the industry will want to assure their customers that the technologies they employ and/or produce are safe, understandable, and accountable. Businesses risk reputational damage, consumer abandonment, regulatory penalties, and financial losses if they fail to incorporate ethical considerations into their overall **governance** frameworks.

Beyond simply being in their self-interest, incorporating tech ethics into the **governance** of businesses and the full lifecycle of their products can enable them to be powerful catalysts for positive societal change. Those leading rather than following can shape the tech sector's potential to be a transformative force for solving humanity's greatest challenges around health, education, food scarcity, environmental sustainability, racial justice, global poverty, and more.

Tara Dawson McGuinness and Hana Schank define "public interest technology" as "the application of design, data, and delivery to advance the public interest and promote the public good in the digital age." When we think about tech ethics in this sense — as a practice rather than a particular product or service — a range of commercial opportunities come into view that extend beyond the narrow ways we typically use the term "technology."

There is a need for public interest technologists devoted to understanding the particular problem one hopes to address by developing a given tech product; conceptualizing a **sociotechnical** approach to address the problem; considering the implications of the proposed approach; curating unbiased, nondiscriminatory, real data sets; centering impacted communities in the design, development, and testing of technologies; assessing risk and ensuring compliance with regulatory requirements; evaluating, monitoring, and auditing technologies while in use; and educating users to ensure understandability and accountability of technologies.

- Opportunities for organizations that can provide services that ensure AI and related algorithmic systems are safe, ethical, aligned, and nondiscriminatory.
- Opportunities for organizations that consult and contract to optimize data privacy, cybersecurity, and ethics in data collection given how fundamental data collection, storage, and security are to the development and deployment of AI systems.
- Opportunities for third-party evaluators that provide independent verification as the United States, European Union, and other regulatory bodies begin to require **risk assessment** and other forms of ethical compliance for high-risk systems.
- Opportunities for tech education organizations (including K-12, higher education, and workforce development) due to the growing need for **sociotechnical** education on how to develop, deploy, work with, and live in a world impacted by AI systems.
- Opportunities for organizations that can provide cybersecurity services, **data ethics** optimization, and assorted technical assistance to civic organizations ranging from religious institutions and community organizations to governmental bodies.
- Opportunities to provide digital literacy and other tech ethical services tailored to specific sectors, especially in urgent public interest areas such as healthcare, law enforcement, public knowledge and cultural heritage institutions, and national defense and security.
- Opportunities for organizations dedicated to designing and leveraging new technologies to address social, political, economic, and environmental problems facing the world.

If we want to build a better world, businesses will have a pivotal part to play. Given this diverse array of opportunities, from conceptualization to deployment and monitoring, tech ethics is not

only vital for mitigating risks and building public trust, but also presents promising avenues for innovation and positive societal impact.

II. CONTEXT

A. Background on the Evolution of Technology Ethics

Technology ethics is, of course, not a new field. It can be traced back far further than there is space to cover here. For a general sense of the landscape today, a brief history might begin in the 1980s and 1990s. Concerns about digital privacy, computer security, and the nascent Internet led scholars, technologists, and their professional organizations to lay the groundwork for the field. Professional associations like the Institute of Electrical and Electronics Engineers (IEEE) and the Association for Computing Machinery (ACM), along with academics like Deborah Johnson, Batya Friedman, James Moor, and Sherry Turkle, helped set guidelines and principles for ethical computing and data science as technologies became more prevalent.

In the early 2000s, the emergence of "civic tech" and "open data" movements added new dimensions to the evolving field. As McGuinness and Schank discuss in Power to the Public, these initiatives were driven by the belief that technology could be harnessed to promote government transparency and citizen engagement. Civic tech advocates employed digital tools to enable public participation in decision-making processes and hold officials accountable. Similarly, the open data movement pushed for public access to government-held information, arguing that making data freely available could spur innovation, improve public services, and empower citizens to engage in data-driven activism.

These developments highlighted the potential for technology to serve as a force for democratic empowerment and social change. They also raised new questions about privacy, security, and the responsible use of public data.

A number of watershed scandals over the past fifteen years have increased public pressure to take tech ethics seriously — particularly around issues like digital surveillance, invasive data collection and breaches of privacy, and **algorithmic bias**. To cite just two prominent examples, Edward Snowden's 2013 revelations about the scale of NSA surveillance programs prompted new scrutiny on technology's role in mass surveillance, while the 2018 Cambridge Analytica scandal raised parallel concerns over data privacy with regard to the power of **Big Tech**. The 2010s and 2020s have been defined, further, by scholars and technologists raising awareness of the dangers of algorithmic systems, and it is not a coincidence that these critical interventions have been led, largely, by women and women of color. Instead, it is indicative of how important it is to center the communities impacted by tech in efforts to make them accountable to the public.

Today, we find ourselves at an inflection point when it comes to tech ethics as a field and how that field intersects with the industry. On the one hand, there are more advocacy organizations, research initiatives, and data activist efforts than ever before. On the other hand, more and more companies are eliminating ethics departments or their equivalents. This shift may favor embedding tech ethics professionals into the company structure. However, the industry currently lacks a systematic sense of what defines a tech ethics job and what are the key metrics for their success. The challenge facing us today is how to translate principles into meaningful change.

B. Rising Concerns Regarding Ethical Practices in the Technology Sector

The field of tech ethics continues to develop against the backdrop of rising concerns, which have only increased in prominence since the release of ChatGPT in 2022. These include: concerns about civilizational and existential risk; concerns about threats to democracy and state interests; concerns about threats to fundamental **human rights**; and concerns about environmental impact and material human costs.

The extent to which AI poses an existential risk to humanity remains **a subject of serious debate.** Moreover, this framing reinforces the asymmetrical power of **Big Tech** companies that use it to justify increased venture capital and state investment, arguing they should be trusted to reign in the potential problems of their own creations — one of the many dangers of what **Evgeny Morozov and others have labeled "technological solutionism."**

This is not to discount serious national and global risks. Yet, while AI poses serious national and global risks — including workplace surveillance, **deepfakes** threatening democratic processes, cybersecurity concerns, and potential proliferation of CBRN weapons — an overemphasis on these geopolitical risks can obscure more insidious impacts. AI systems often amplify existing biases and inequalities, violate privacy rights, and disproportionately harm vulnerable communities who are rarely centered in their development.

Ethical concerns extend to the **governance** of tech companies themselves, including issues of power concentration, lack of accountability, and "**ethics washing**." The dominance of a few **Big Tech** companies, largely led by white, cisgender male technologists, raises concerns about the equitable distribution and **governance** of AI capabilities.

Moreover, the material and environmental impacts of AI development are significant. Training and operating **large language models** contributes to climate change through massive energy consumption and e-waste production. The AI supply chain also involves exploitative labor practices, with precarious workers often performing essential "**ghost work**" under poor conditions and with limited rights.

When the concerns and experiences of the Global South and marginalized communities across the world are centered, other issues emerge that must be attended to if PIT efforts are to be successful. These include fair labor practices, for example, given the ways in which outsourced labor that lacks systems of accountability to prevent exploitation is a rampant force that powers AI advances. This includes the tremendous energy costs and environmental impact required to develop and run powerful foundation models, which has an exorbitant detrimental impact on the earth in a moment of climate catastrophe. This also includes the digital inequity that defines our current context as well, where vast swaths of humanity do not have the electricity, internet access, or fundamental resources required to partake in the benefits of AI.

C. Emergent Regulatory Landscape and Its Impact on Technology Ethics

These concerns are among the factors contributing to an emergent regulatory landscape. It remains unclear which frameworks will be adopted and successfully enforced, or whether and how they will operate in relation to one another in a global economy. Nevertheless, regulations will be one of the most significant drivers of tech ethical commercial opportunities in the coming years. There is, as of yet, no consensus on whether and to what extent regulations inhibit innovation. Of course, **Big Tech** companies competing for market dominance have different motivations in answering this question than the governing bodies hoping to regulate them, not to mention the activist and advocacy groups struggling to hold them accountable. Many questions remain unanswered and will certainly be tested in the coming years. For instance, will creative applications of existing laws (e.g. privacy, copyright, and anti-discrimination laws) offer sufficient protections, or, will new technologies require new jurisprudence? Differing emphases notwithstanding, the **European Union's AI Act** (**EU AI Act**), President Biden's **Executive Order** (AI EO), and the **Cyberspace Administration of China** (CAC)'s interim measures on GenAI services all give a sense of how state actors will attempt to carve a middle path to ensure AI systems do no harm while allowing space for technological innovation.

The AI EO's thesis statement may be that "artificial intelligence (AI) holds extraordinary potential for both promise and peril," yet protecting against peril is its overriding concern. Perils identified in the text range from the sociopolitical (i.e. AI systems that reinforce discriminatory practices in policing, housing, healthcare, and education) to the geopolitical (i.e. systems that disseminate disinformation, or lower the barrier of entry to the production of chemical, biological, radiological, or nuclear weapons), but its emphasis is on the latter, the potential threats to U.S. national interests and on ensuring the U.S. maintains its cutting edge in the

emergent market. The AI EO articulates the need for third-party evaluation and regulation of dual-use foundation models that pose these potential risks, yet, as an executive order (rather than an act of legislation), it remains unclear whether and to what extent it will be enforceable in the coming years. The AI EO does aim to promote the "promise" of AI systems by incentivizing the development and deployment of particular kinds of AI systems (e.g. technologies that might offer promising benefits to democracy, or that will offer significant benefits to American health and prosperity) while also ensuring that the US is competitive and, indeed, remains the global leader in AI.

More recently on the legislative level, in May 2024, Senate Majority Leader Chuck Schumer released **a legislative roadmap for AI policy,** following a year-long series of closed-door "AI Insight Forums." While the roadmap pledges $32 billion for AI research and development, it notably lacks concrete regulatory guardrails to protect civil and **human rights**. In response, a coalition of civil society organizations published **"Put the Public in the Driver's Seat: Shadow Report to the US Senate AI Policy Roadmap."** This report highlights the stark contrast between global regulatory action and U.S. Congressional inaction, despite a strong public mandate for AI regulation. The Shadow Report identifies 11 key areas requiring enforceable laws that protect against the dangers posed by AI systems. These include racial justice and equity, immigration, AI accountability, labor protections, privacy and surveillance, competition policy, consumer protection, safeguarding democracy, public interest-focused industrial policy, poverty prevention, and climate change considerations. Taken together, Schumer's roadmap and the Shadow Report's critique provide insight into the complex and contested landscape of emerging AI regulation in the United States, highlighting the tension between industry influence and public interest advocacy in shaping policy.

The EU AI Act, **which was officially adopted by the EU Parliament on March 13, 2024,** also attempts to strike a balance between incentivizing innovation while preventing harms. Yet, it is more explicit that the AI EO in its attempt to protect against **automated decision-making** systems that might violate "fundamental rights" of "natural persons." Whereas the AI EO focuses the bulk of its attention on dual-use foundation models that might pose national security risks (CBRNs, disinformation, synthetic biology), the EU AI act considers anything that has the potential to adversely impact the livelihoods of natural persons to be "high risk." This includes systems that pose risks on a national and international scale to be sure, but just as much so those that threaten individuals and communities. An AI system used in credit scoring, or in hiring, or in education admissions processes could all significantly impact people's lives and, thus, are considered high risk. Any system deemed "high risk" (and some deemed "medium risk") will be required to demonstrate compliance before entering the market; maintain documentation and **transparency** that ensures they are continuing to comply with said Regulation; submit to compliance monitoring and assessment on an ongoing basis; and be prepared to temporarily come off the market to address non-compliance, or face significant fines.

China is another regulatory context that is essential for any truly global understanding of the tech ethics landscape. The **Cyberspace Administration of China** (CAC) released their **"Interim Measures for the Management of Generative Artificial Intelligence Services"** in advance of either the AI EO or the **EU AI Act**. In contrast with the **EU AI Act** which aims to be a comprehensive regulation for all AI applications from the outset, **China has issued targeted regulations for specific applications** like recommendation algorithms, synthetically generated content, and **generative AI** models, before attempting a more comprehensive AI law. For example, the

regulation on recommendation algorithms requires providers to register detailed information about their algorithms and training data in a government registry. The Chinese approach also places a strong emphasis on information control and aligning AI systems with state priorities, as seen in requirements that AI-generated content must "embody Core Socialist Values" — a focus that contrasts with the emphasis on mitigating risks like bias and privacy violations in the US approach.

In addition to regulations imposed at the national and international level, various U.S. states, municipalities, and agencies will also begin to enforce their own requirements as well. For instance, the **Federal Communications Commission** (FCC) unanimously adopted a ruling that considers robocalls using AI-generated voices as "artificial" voices under the federal Telephone Consumer Protection Act, effectively making their use illegal. **A number of U.S. states have begun work on AI-related legislation**, which includes funding for AI research, data privacy protections, preventing **algorithmic bias**, challenging **deepfakes** and disinformation, and more. Local laws, like **NYC Local Law 144**, are emerging too. While there are questions around the extent to which NYC 144 has or can be successfully enforced, it is a regulation that addresses the use of **Artificial Intelligence** (AI) in hiring and employment decisions within New York City. The law took effect on July 5, 2023, and applies to employers and employment agencies that use Automated Employment Decision Tools (AEDTs) for making employment-related decisions, such as hiring, firing, and promotions. Other municipalities and state governments will likely pass their own legislation regulating AI systems in the coming years.

Taken together, these regulations will increase market demand for tech ethics services and products as well as structure the conditions

under which tech ethics continues to develop as a field. There are, admittedly, drawbacks to placing too much emphasis on state actors and regulatory bodies in tech ethical considerations. Nevertheless, there is good reason to remain skeptical that businesses, **Big Tech** in particular, would voluntarily submit to the necessary regulation, or design systems with adequate ethical and/or socially just inputs and outputs without external pressure, especially in our current **digital neoliberal** context. To cite just one example, **Apple shareholders' recently rejected a proposal** from the AFL-CIO seeking greater **transparency** on how the company uses AI and what ethical guidelines it employs to navigate its use. The emergence of regulatory frameworks mark an inflection point insofar as they represent one important pressure point on the tech industry.

D. Purpose and Scope of the White Paper

This white paper aims to assess the current state of the **technology ethics** field. Its primary purpose is to identify needs, demands, and opportunities for the commercialization of tech ethics services and products. The scope of the paper encompasses a wide range of areas, including **algorithmic auditing**, third-party **vetting** of AI software, **data ethics** consulting, compliance with emerging regulatory frameworks, technology ethics education and training, and specialized consulting services tailored to various sectors such as healthcare, law enforcement, and national security. The paper offers real-world case studies and success stories, highlights challenges and mitigation strategies, and explores future trends and innovations that may shape the evolving **technology ethics** landscape. By delivering a thorough analysis and actionable insights, the aspiration is that this white paper encourages and empowers diverse stakeholders to navigate the complexities of the tech ethics landscape and successfully bring to market products and services that can build a more just and equitable world for all.

III. OPPORTUNITIES FOR COMMERCIALIZING TECHNOLOGY ETHICS SERVICES AND PRODUCTS

The need is clear. As we integrate machine-learning, algorithmic, and other AI systems into more and more facets of human life, it will be increasingly urgent that we ensure those systems be aligned not only with human intentions but with values oriented toward a more just world for all. A panoply of services will be necessary to protect against the risks, harms, and dangers posed by these systems and the ends to which they are employed. There is also a growing sense of the creative potential AI systems and other emergent technologies hold for addressing pressing problems in the public interest. AI systems operate at a speed and scale that could prove invaluable to human efforts to detect diseases, develop new medical treatments, prevent climate catastrophes, increase access to a range of essential social services, and more. Moreover, the regulatory requirements and incentive structures coming into effect in China, the European Union, the United Kingdom, the United States, and other contexts create market demand for a range of commercial opportunities. What follows is a survey of those opportunities organized by A) the tech ethics products and services for which there is market demand, and B) the sectors that have a distinctive set of tech ethics needs.

A. Products and Services

1. Algorithmic Auditing

Algorithms are essential to modern life yet largely invisible to those impacted by them. They inform decisions across society in ways incomprehensible to most people, often operating as "black boxes"

due to their opacity and complexity. Pioneering research has shown that algorithmic systems can be not just inaccurate, but actively biased, reinforcing various forms of discrimination. This necessitates algorithmic audits - systematic examinations of these systems to ensure accountability, **transparency**, and ethical standards. These audits scrutinize a model's entire lifecycle, assessing factors like optimization criteria, data inputs, and feedback loops to identify and mitigate risks and biases.

The **algorithmic auditing** ecosystem is evolving, with organizations like the **International Association of Algorithmic Auditors** (IAAA) leading the way. Algorithmic audits aim to identify, address, and correct harms across various domains, employing a multidisciplinary approach that includes **governance**, empirical, and technical audits. This reflects the **sociotechnical** nature of algorithmic systems.

Market demand for **algorithmic auditing** is growing due to increasing regulation and public awareness. The AI Executive Order and **EU AI Act** both call for evaluation and auditing of AI systems, particularly those posing risks to national security, civil rights, or fundamental rights. These regulations, along with local laws like NYC Local Law 144, often require third-party evaluations or conformity assessments for medium- and **high-risk AI systems**.

Nonprofit organizations and research institutes have paved the way in the field. The **Algorithmic Justice League,** founded by Joy Buolamwini, combines art and research to highlight the social implications of algorithms, **artificial intelligence**, and technology. They work “to raise public awareness about the impacts of AI, equip advocates with resources to bolster campaigns, build the voice and choice of the most impacted communities, and galvanize

researchers, policymakers, and industry practitioners to prevent AI harms."

There are already some successful efforts at commercializing algorithmic auditing. O'Neill Risk Consulting & **Algorithmic Auditing (ORCAA)**, founded by data scientist Cathy O'Neill, has an interdisciplinary staff whose work includes **algorithmic auditing**, quantitative testing for regulatory compliance, AI **governance** and risk management consultation, "cockpit design," and education and training. **Luminos Law** leverages its legal expertise to provide consultancy services focused on developing robust compliance strategies and ensuring adherence to relevant regulations and guidelines. **BABL AI** is another early exemplar in the field, founded by Shea Brown and co-founded by Jovanna Davidovic, Jeffery Wrecker, and Khao Lam. BABL AI's primary work is algorithmic and AI systems auditing, tailored to meet the requirements of specific regulations (i.e. **EU AI Act** Conformity Assessment Readiness audit; NYC Bias audit; NIST Risk Management Framework Readiness audit; Digital Services Act audit; EEOC AI Bias audit).

2. Third-Party Vetting and Verification of AI Software

Algorithmic auditing is not a panacea, though, often failing to detect biases evident to users after deployment. As AI integration grows, there's increasing demand for third-party evaluators to vet AI systems before procurement, ensuring **alignment** with organizational values. The **National AI Research and Development Strategic Plan** (R&D Strat Plan) emphasizes the need for trustworthy, interpretable AI that's verifiable and certifiable, going beyond safety and **transparency** to include privacy protection and bias avoidance.

Regulatory frameworks like the **EU AI Act and the AI Executive Order** require independent third-party assessments for **high-risk AI systems**. These assessments involve continuous, iterative processes throughout the system's lifecycle. The AI EO also calls for international cooperation in developing AI standards, emphasizing trustworthiness, verification, and assurance. Industry-wide standards and regulatory requirements are crucial for effective **vetting** and verification, with organizations like the IEEE Standards Association playing a key role.

The demand for AI **vetting** extends beyond regulatory compliance. The "**black box**" nature of AI systems is a major concern for users, necessitating services to navigate this complexity. Organizations across various sectors will need assurance that AI systems align with their values and mission. Civic institutions, community organizations, law firms, and universities will require formal verification that AI systems support rather than undermine their objectives.

3. Consulting for Best Practices in Data Ethics

In tandem with and in addition to the need to audit and vet AI systems, the market demand for **data ethics** consultation services is growing across diverse industries that can be divided into at least two different contexts: **data ethics** research and consulting services for the training of high risk AI systems, and consulting for best practices in **data ethics** for industries that employ AI systems to collect, retrieve, and store personal data.

To begin with the former, AI systems are only as ethical and accurate as their training data. Biased or limited datasets produce biased AI outputs. This creates an opportunity for commercialization in research and consulting services for ethically-aligned,

community-centered data curation. **Emerging regulations**, such as the AI EO and **EU AI Act**, emphasize the importance of representative datasets, equity principles, and proper data **governance**. These regulations will require third-party evaluators to assess data collections for potential risks to fundamental rights, creating additional demand for specialized services.

Beyond AI training**,** there's a growing need for **data ethics** consulting in industries handling large amounts of user data. The rise of **machine learning** systems has intensified concerns around privacy, security**, transparency,** and ownership. **Data ethics** services are crucial to ensure ethical data collection practices, user consent, data disaggregation, and protection of personal information. As highlighted Jennifer King and Caroline Meinhardt's white paper for **Human-Centered AI** at Stanford, **"Rethinking Privacy in the AI Era: Policy Provocations for a Data-Centric World,"** the shift to **generative AI** presents new challenges in **data ethics** that can't be addressed solely through individual data rights. This evolving landscape will drive increased demand for **data ethics** consultation services in the coming years.

4. Workforce Development, Professional Development, and AI Ethics Educational Services

Tech education broadly, and digital literacy in particular, is now an ethical, economic, and democratic imperative. The immediate focus for many policymakers is on workforce development in anticipation of widespread AI integration and automation. This will undoubtedly be a key area of opportunity. There will also be a need for AI literacy at virtually every level: from K-12 to higher education to community education and ongoing professional development. If AI systems are to contribute to human flourishing and help us create a more just

and equitable world, people will need to have a basic sociotechnical understanding of (1) what they are and how they operate; (2) how to engage them critically and use them ethically; and (3) the material costs, potential consequences, and broader implications of these technologies. This will require interdisciplinary expertise. As **Suresh Venkatasubramanian** has argued, "effective AI work is really **sociotechnical** rather than purely technical. This requires more STEM education for students and researchers in the social sciences and humanities, and broadening the diversity of perspectives in the field of AI broadly speaking." There will, hence, be extensive opportunities for curricular development and the delivery of educational services.

There is a need for AI education and training for K-12 educators looking to integrate new technologies into their classrooms; for higher education faculty and administrators seeking to deploy AI systems, not just in the classroom but to aid in admissions processes and other administrative duties; for the broader American workforce, as employers look to integrate AI into the workplace and employees look to augment their work with the help of AI; for diverse professionals recruited into the tech sector; for communities, particularly minoritized and otherwise vulnerable communities, and civic institutions looking to deploy AI systems to improve things ranging from infrastructure to town hall meetings. There is an opportunity here for tech ethics organizations to position themselves as leaders in this field, especially if they leverage their values, diversity, and interdisciplinarity as assets. This is especially true given the fact that higher education has already begun integrating AI systems into the life of the university. For example, **Arizona State University recently entered into a partnership with OpenAI** to integrate ChatGPT into academic and student life. Given the market demands on universities to compete in a context of increasing budget austerity and shrinking student populations, ASU will certainly be just the first of many colleges and universities to do so,

further highlighting the tech ethics commercialization opportunities to ensure the ethical application of AI systems not just on the level of the companies producing them but also the institutions employing them.

There are efforts already underway to incentivize AI ethics education. The AI EO calls on the Secretary of Education to develop resources, policies, and guidance regarding AI in education that will "address safe, responsible, and nondiscriminatory uses of AI in education" and will include "appropriate human review of AI decisions, designing AI systems to enhance trust and safety, and align with privacy-related laws and regulations in the educational context, and developing education-specific guardrails." There is explicit emphasis on the impact AI systems might have on "vulnerable and underserved communities," so this is an important area where tech ethics organizations are essential. **The R&D Strat Plan** calls for federal investment in the development of instructional material and the support of higher education staff in this area:

High-quality, domain-specific, and appropriately challenging lessons are needed for introducing students to critical thinking skills that will help them understand and evaluate AI systems. The research required to properly identify and curate the right content for a given area and level of study requires considerable effort. Further research is needed to sort out the best pedagogy and media through which to convey this content, as well as to identify and curate best practices for training instructors. [...] Additionally, it is important that any AI materials, training programs, or systems are accessible, equitably promulgated, and broadly representative, especially given current inequalities among students' and educators' access to resources.

The emphasis here on accessibility and equity as well as the call to cultivate critical thinking skills points to an interdisciplinary approach to AI education that include but extend beyond STEM into the social sciences, arts, and humanities.

5. Technology Ethics Consulting and Security Services for Civic Institutions and Community Organizations

Community organizations and civic institutions are increasingly in need of tech ethics guidance and services, particularly regarding data privacy and cybersecurity. We have already noted how AI systems pose special risks to marginalized and vulnerable populations. The disproportionate adverse impact AI can and will have on these communities will require specialized community education as well as services geared toward civic institutions and community organizations dedicated to working in and serving these communities, particularly given the fact that such institutions and organizations in the public sector often lack the requisite resources and technical expertise to do so themselves. Cybersecurity services, for instance, are relevant for groups ranging from religious communities and schools to congregation-based community organizations, all of which may benefit from consulting to help navigate procurement of services.

The AI EO, **EU AI Act**, and countless studies on the societal impact of AI have identified protecting minoritized, underserved and vulnerable communities from potential AI risks and harms as a priority. This presents potential commercial opportunities for tech ethics companies that can provide advocacy, outreach, community education, and serve as conveners connecting community members with public interest technologists. There will be a premium

here, on companies that explicitly build and emphasize commitments to diversity, equity, and inclusion, **social justice,** and the public interest in their business model. The AI EO cites specific examples where such opportunities might arise, such as promoting notice to recipients about the use of AI systems in policing practices, regular evaluation of systems to detect unjust denials, and community engagement around potential ableist assumptions built into AI systems.

Tech ethics companies can offer essential services to community organizations beyond education and liaising as well. Given the data collection required for community organizing, these organizations could benefit from data optimization and training in the responsible use of AI and algorithmic systems to enhance their social impact. The AI EO emphasizes the importance of aligning AI systems with not just human intentions but also the needs of community stakeholders and vulnerable populations. This is a space for community-centered commercial opportunities that may arise since external, third-party oversight will be crucial to making this oversight a reality. The EO specifically notes that, to further the protection of civil rights, agencies "shall also consider opportunities to increase coordination, communication, and engagement about AI as appropriate with community-based organizations; civil-rights and civil-liberties organizations; academic institutions." Facilitating connections between the tech sector and community stakeholders, thus, will be an important opportunity space for public interest technologists and ethical tech companies.

6. Social Justice Tech Ecosystem

Given the innumerable ethical issues raised by emergent technologies and the complexity of the questions presented by their widespread adoption, the problem space is so vast that there is not

only a need for all the products and services listed above, but also for the coordination and collaboration across a social justice tech ecosystem. Organizations dedicated to fostering such an ecosystem could offer a range of products and services aimed at connecting, empowering, and standardizing practices within the field. At the core of this model could be a membership-based structure, similar to that of the IEEE, but tailored specifically to the needs of public interest technology (PIT) and social justice tech practitioners. This membership could provide access to a network of like-minded professionals, organizations, and resources, creating a virtual and physical space for collaboration, knowledge-sharing, and collective action.

Key offerings within this ecosystem might include certification programs for PIT and ethics-focused businesses, helping to establish credibility and standards within the field. These certifications could be complemented by the development of Key Performance Indicators (KPIs) for PIT and ethics-related jobs, providing a framework for assessing impact and effectiveness across the sector. To address the often project-based nature of PIT work, the organization could facilitate a labor network, allowing member organizations to share workers and expertise as needed. This could help create more stable employment opportunities while ensuring that a diverse range of projects have access to skilled professionals.

Beyond these structural supports, the organization could offer a suite of services designed to enhance the impact and sustainability of **social justice** tech initiatives. This might include coordinating and collaborating on third-party evaluations and audits, developing rubrics for assessing the potential impact of PIT projects, and facilitating cross-sector collaboration between academia, industry, and government to translate research into practical applications.

The organization could also serve as a collective voice for advocacy on PIT issues, leveraging the combined influence of its members to shape policy and public discourse. Finally, it could play a crucial role in connecting PIT enterprises with potential funders, helping to ensure the financial sustainability of impactful projects and organizations within the ecosystem.

B. Sectors

1. Small and Medium Businesses (SMBs)

Whether it be on the level of policymakers, pundits, or popular opinion, tech ethics discussions tend to dedicate the bulk of their attention to whether and how to regulate **Big Tech**: Alphabet, Amazon, Apple, Meta, and Microsoft. This is for good and obvious reasons. These companies are in the business of funding and developing many of the dual-use foundation models that pose the greatest potential risks to society. They also wield exorbitant power due to not only their wealth and influence, but also their collection and control of unparalleled amounts of user data. Nevertheless, AI systems present a host of potential benefits and risks for small and medium entrepreneurs and businesses (SMBs). SMBs will, thus, be in the market for tech ethics services and products to help them navigate our current landscape.

AI and other automated technologies present opportunities for SMBs to optimize their impact in ways they otherwise might not be able to afford given their inherent resource limitations. For example, if a start-up cannot afford to hire a communications team, they may use GenAI to help produce internal and external comms content. However, without technical or techno-ethical expertise, SMBs may rely on AI without considering the potential unintended

consequences of doing so as well as the serious legal liabilities that might present themselves. This extends beyond AI to the area of **data ethics** as well. For example, a **2019 survey of 716 small business leaders across Europe** revealed widespread ignorance about data security tools and only loose adherence to key privacy requirements of the GDPR, despite substantial investments by many businesses to achieve compliance with the regulation. There will, thus, be a need for tech ethics services and products tailored for SMBs in particular. This will include consideration of the viability of proposed technical solutions, consultation on the **governance** practices required for incorporating technologies into the company's workflow, **vetting** potential systems prior to procurement, instructional services on how to use said systems responsibly and ethically, etc.

The AI EO has created an incentive structure to specifically support SMBs. The Small Business Administration (SBA) has been instructed to "prioritize the allocation of up to $2 million in Growth Accelerator Fund Competition bonus prize funds for accelerators that support the incorporation or expansion of AI-related curricula, training, and technical assistance, or other AI-related resources within their programming. The AI EO also specifically instructs the SBA to conduct outreach and raise awareness for small businesses that "serve or with experience serving underserved communities" to use capital-access programs for AI-related purposes."

2. Public Knowledge and Cultural Heritage Institutions

AI and **machine learning** offer exciting opportunities to enhance archival programs and broaden access to content. However, their implementation must address not only AI bias but also existing institutional biases in curation and cataloging. The historically limited diversity in library and archives professions has contributed to discriminatory practices that could be exacerbated by uncritical

AI adoption. Careful consideration of tech ethics principles is crucial when applying these technologies to public knowledge and memory institutions.

While mass digitization and AI can extract new information from archives, potentially reducing biased descriptions, care must be taken not to inadvertently rewrite history. AI applied to historical records could uncover evidence of systemic issues, such as racism in government actions. However, it's crucial to ensure AI systems don't perpetuate historical biases. **Transparency** in data analysis, bias identification, and correction efforts is essential.

Adopting AI technologies in archival institutions requires significant infrastructure investment, including IT resources, GPUs for deep learning, increased storage, and network capacity. Cybersecurity measures are critical, as are staff training in AI risk management. Given AI's growing role in accessing public knowledge, systematic auditing of AI software for public archives is imperative to address potential biases and ensure the integrity of our cultural heritage.

Public knowledge institutions like universities, libraries, and museums thus present a viable market for tech ethics companies. These organizations can benefit from tailored products and services such as educational curricula, training programs, curated resources, and consultancy services. For universities, these offerings can enhance AI literacy and ethics awareness across disciplines. Libraries can provide curated collections and public workshops, while museums can offer immersive exhibits on technology's societal impact. Tech ethics companies can also help these institutions develop their own ethical frameworks for technology use. By aligning with these institutions' missions, tech ethics companies can disseminate ethical principles while tapping into a socially conscious market segment.

3. Federal Agencies and Regulatory Bodies

The commercialization of tech ethics services and products extends beyond the private sector, with significant opportunities in government. As McGuinness and Schank argue, "governments and nonprofits must adapt to the modern world and find a way to deliver for the people." The AI Executive Order (EO) addresses this need by encouraging federal agencies to establish guidelines for AI use rather than imposing broad bans. It requires each agency to designate a Chief Artificial Intelligence Officer and aims to increase AI talent in government through targeted recruitment and hiring initiatives.

The EO also establishes the White House AI Council, comprising leaders from a wide range of federal departments and agencies. This council represents the breadth of government entities that will require tech ethics expertise, including but not limited to:

- Key cabinet departments (State, Treasury, Defense, Justice, Labor, etc.)
- Agencies focused on science, technology, and economic policy (NSF, OSTP, OMB)
- National security and intelligence agencies
- Regulatory bodies (FTC, CFPB, NIST)

This diverse representation underscores the pervasive need for AI expertise and ethical considerations across all areas of government, creating significant market potential for tech ethics services in the public sector.

4. Health and Human Services

Healthcare is a sector where AI presents both exciting potential and significant risks. AI can assist in early disease detection, drug discovery, personalized medicine, and efficient healthcare delivery. It can analyze medical images, patient data, and scientific literature to identify patterns, make diagnoses, and suggest treatments. AI systems could also serve as virtual assistants, easing the burden on physicians overwhelmed by expanding medical knowledge and bureaucracy.

The AI Executive Order recognizes these benefits and establishes incentives for AI talent in healthcare. For example, it instructs the HHS Secretary to prioritize grants supporting **responsible AI** development, including AI-enabled tools for personalized immune-response profiles and initiatives improving healthcare-data quality for AI tools in clinical care and public health.

However, these possibilities raise ethical concerns. Data privacy, security, and patient confidentiality are paramount, given AI's capacity to handle massive amounts of sensitive health data. There's also a risk of exacerbating healthcare inequalities if AI adoption is uneven or systems are trained on biased data. To address these issues, the AI EO mandates an HHS AI Task Force to develop policies for **responsible AI** deployment in health services. Furthermore, AI integration in healthcare raises broader questions about balancing efficiency with core professional values like deep listening and holistic care. While not mutually exclusive, incorporating AI necessitates critical reflection, highlighting the need for AI ethics education and **governance** in healthcare technology.

5. Law Enforcement

Commercializing tech ethics services and products for law enforcement is of paramount public interest. Law enforcement agencies already employ a range of drone, surveillance, algorithmic, and other AI-powered technologies to aid in policing practices. Software and technology companies regularly develop and market cutting-edge products for law enforcement, often without consideration of the broader justice and equity concerns they might raise. Police departments across the United States, for example, have used algorithmic "**predictive policing**" programs that have been found to disproportionately target Black and Brown communities and contribute to over-policing minoritized communities. **The U.S. Government Accountability Office** found that a number of federal law enforcement agencies employed **facial recognition** technologies with no training requirements or policies to protect civil liberties. Even seemingly innocuous or potentially beneficial technologies, like automated license plate readers and body cameras, raise fundamental concerns about privacy, data security, and consent.

It is not a question of whether law enforcement agencies will procure these technologies. Instead, it is incumbent upon tech ethics organizations to effectively commercialize (1) the design of ethically-aligned tools, systems, and datasets; (2) the development of tech products compliant with civil and **human rights** that can aid in efforts to promote public safety; (3) third-party **vetting** services to aid agencies in the procurement of technologies in the public interest; (4) consulting services to provide training in ethical use of technology and education on the potential problems and dangers of algorithmic systems; (5) the monitoring of AI-augmented law enforcement practices; and (6) public advocacy to inform citizens of the use and their fundamental rights in relation to these technologies.

There are already **social justice** technologists working to address these concerns. For instance, in the United States, a number of technologists of color politically activated by the Black Lives Matter uprisings founded **JusticeText** in 2015. Their company developed software that responds to the challenge of processing large amounts of video evidence to support fair legal processes. Their work promotes legal accountability by helping public defenders efficiently analyze video data from police interactions. **Measure** is another **social justice** tech initiative born of the energy of Black Lives Matter mobilization. This Black women-led non-profit organization builds capacity within marginalized communities to address their most pressing needs by elevating lived experiences as data and training community members as data activists. Finally, in Hyderabad, India, the **Criminal Justice and Police Accountability Project** (CPA Project) demonstrates how tech-enabled research and advocacy can challenge systemic biases in policing. Founded in 2020, the CPA Project uses data collection and analysis to document and combat the disproportionate targeting of marginalized communities, particularly Denotified Tribal communities, by the criminal justice system.

Regulatory frameworks are making tech ethics services for law enforcement not just a public need, but a legal requirement. The AI Executive Order mandates improved training for federal law enforcement on AI-related cases involving civil rights violations. The **EU AI Act** goes further, prohibiting real-time biometric identification in public spaces for law enforcement except in limited cases. It also classifies AI systems that might discriminate or unjustly single out individuals as "high-risk," requiring them to undergo conformity assessments. These regulations emphasize the need for high-quality data, robust design, and **transparency** in AI systems used in law enforcement to protect fundamental rights, including the right to a fair trial and presumption of innocence.

IVCASE STUDIES AND SUCCESS STORIES

The problem space is immense. The need could not be more urgent. Activist and advocacy collectives, academic research institutes, and other not-for-profit organizations continue to do pioneering work unveiling the unjust systems behind and consequences of technologies as well as promoting tech ethical best practices. The commercialization of tech ethics services and products can contribute to the transformative work of these nonprofit organizations by bringing the principles and practices of tech ethics, public interest tech, and **social justice** tech to market. This may prove essential given the inherent limits and challenges of donation-based and grant-driven work. The solution space for commercializing tech ethics is still emerging, yet there are real success stories in the tech sector as well as useful analogues in other sectors that can serve as potential models moving forward. It is not wishful thinking to imagine entrepreneurs who seek to elevate safety and ethics while also generating revenue by commercializing services and products. There are companies and organizations doing so right now.

ABusiness Structure

When considering the commercialization of tech ethics services and products, it's crucial to examine various business structures that align organizational form with social impact goals. Traditional corporate models often prioritize profit over social good, which can be at odds with the mission of public interest technology. Alternative business structures such as **public benefit corporations (PBCs)** and cooperatives can offer innovative ways to balance financial sustainability with ethical commitments and social responsibility.

1. Public Benefit Companies

If the ultimate goal is to commercialize tech products and services that serve the public interest and contribute to socially just outcomes, the public benefit company (PBC) is an especially apt framework to consider. A PBC is a legal corporate structure that obligates the company to consider the interests of multiple stakeholders, not just shareholders but also employees, customers, and the public writ large. This can make PBCs attractive to mission-aligned investors and contribute to a longer-term, sustainable approach to value creation. Elon Musk's hostile takeover of Twitter in 2022 serves as a recent counterexample that highlights the potential benefits of a PBC. As a traditional corporation, Twitter's board had a fiduciary duty to consider Musk's takeover bid and make a decision that maximized shareholder value. If Twitter had been a PBC, on the other hand, the board would have had a legal obligation to consider the impact of the takeover not only on potential profit margins for shareholders but also on the public benefits the platform provided, which might have led to a different outcome. While a PBC structure does not guarantee protection against bad actors or hostile takeovers, it does offer a means for companies to align their legal form with their mission-driven approach.

One of the earliest examples of a technology company adopting the PBC structure to commercialize tech ethics is **Coursera,** an online learning platform founded in 2012 by Stanford University professors Andrew Ng and Daphne Koller. The company partners with universities and organizations to offer a diverse range of online courses, specializations, and degree programs, empowering individuals to gain valuable skills and knowledge at their own pace. By leveraging tech to democratize education, making it more accessible and affordable, Coursera demonstrates the potential for PBCs to create both social impact and commercial success.

In the biotechnology sector, **United Therapeutics Corporation** became the first publicly traded biotech company to convert to a PBC in 2021. The company's groundbreaking work in xenotransplantation, which involves genetically modifying animal organs for human transplantation, has the potential to address the critical shortage of donor organs and save countless lives. By focusing on developing innovative treatments for rare and life-threatening conditions, such as pulmonary arterial hypertension, United Therapeutics exemplifies how a biotech PBC can prioritize patient well-being while creating sustainable value for shareholders.

Greenlight Biosciences, founded in 2008, is another biotech company that has embraced the PBC structure to advance its mission of creating "sustainable solutions for a growing world." Greenlight Biosciences leverages its proprietary RNA production platform to develop innovative solutions for agriculture, human health, and industrial biotechnology. By applying its technology to address critical issues like food security, infectious diseases, and climate change, Greenlight Biosciences demonstrates the potential for biotech PBCs to create value for a wide range of stakeholders while driving positive social and environmental impact.

These three companies, spanning education and biotechnology, in addition to AI research like Anthropic, demonstrate a growing trend of tech companies adopting the PBC structure to prioritize social responsibility alongside financial success.

2. Labor Cooperatives

In addition to the legal framework of a public benefit corporation, labor cooperatives offer a compelling alternate model for organizations seeking to align their business structure with their

social impact goals, particularly in addressing the significant labor issues that have arisen in the tech sector in recent years. Though they are often sidelined in tech ethics conversations, particularly those that center the concerns of the Global North, critical consideration of human costs, environmental impacts, and fair labor practices when developing and auditing AI systems is an urgent tech ethical issue. Cooperatives, owned and democratically controlled by their members, can prioritize fair labor practices and worker empowerment in ways that traditional corporate structures may not. They also, by their very nature, center the communities and contexts most impacted by the tech development and deployment in question.

In the tech ethics space, labor cooperatives like **Sassafras Tech Collective** and **Research Action Design** (RAD) demonstrate how cooperative principles can be applied to software development, prioritizing fair labor practices and worker empowerment. Platform cooperatives, such as those supported by the **Platform Cooperative Consortium**, offer alternatives to traditional gig economy platforms, allowing workers to collectively own and govern the platforms they use. For instance, driver-owned ride-hailing cooperatives could provide more equitable working conditions compared to corporate-owned alternatives. While not a tech cooperative, the **Self Employed Women's Association** (SEWA) in India shows how cooperatives can empower marginalized workers on a large scale, providing a model that could be adapted for the tech sector.

Cooperatives can address several key issues in the tech ethics landscape. They can ensure fair compensation and working conditions for often-invisible labor in AI development, such as data labeling and content moderation. By giving workers a voice in decision-making, cooperatives can help ensure that ethical

considerations are integrated throughout the development process. Further, the cooperative model aligns well with the need for t**ransparency** and accountability in AI systems, as democratic **governance** can lead to more open and responsible practices.

While cooperatives face unique challenges in scaling and accessing capital, they present a promising option for tech ethics ventures aiming to address labor issues, ensure fair compensation, and embody principles of democratic **governance**. By incorporating cooperative models, organizations in the tech ethics space can potentially transform labor practices and create more equitable economic structures. As the field of tech ethics continues to evolve, cooperatives could play a crucial role in ensuring that the benefits of technological advancement are more equitably distributed among workers and society at large.

B. Organizations

1. Tech Products that Serve the Public Interest

Developing technologies that serve the public interest is another approach to commercializing tech ethics products and services. Whether or not the founders of these startup companies and tech organizations conceive of themselves as "public interest technologists," they exemplify how commercial tech products and services can be developed that help build a better world.

- **Cortico** is a nonprofit organization that develops technology to foster constructive public dialogue. Their main product, the Local Voices Network, is an AI-powered platform that combines in-person community conversations with digital tools to surface patterns, insights, and areas of common ground and differing perspectives on important issues. By enabling ethical and inclusive communication, Cortico's platforms foster constructive dialogue and address issues such as online harassment and misinformation.
- **Flikshop** is a company, founded by a formerly incarcerated man, that helps incarcerated individuals stay connected with their families through a digital platform that allows users to send photos and personal messages as postcards to loved ones in over 2,000 correctional facilities across the US. The goal is to maintain family connections and support reentry planning. This is an especially important platform (developed by a formerly incarcerated person), given how predatory the most prominent tech providers in prisons have been historically.
- **JusticeText** is a software company that focuses on promoting accountability in the legal system by helping public defenders efficiently analyze video data from police interactions. Founded by technologists of color politically activated by the Black Lives Matter uprisings that began in 2015, they are addressing the challenge of processing large amounts of video evidence to support fair legal proceedings.
- **MABLE** is a software company that offers tech solutions to streamline the public housing application process. Their platform provides user-friendly tools for both housing specialists and applicants, including a comprehensive digital form for compliance-related data collection, AI-powered document verification, automated scheduling, and email notifications. MABLE's system is accessible across devices and includes

analytics for operational insights. By simplifying and automating key aspects of housing applications, MABLE aims to reduce stress, save time, and improve access to housing services.

- **Nucleos** is a public benefit company that provides a secure online learning platform for incarcerated individuals to access educational content, communicate with instructors, and earn credentials. The platform is designed to work within the constraints of prison security systems while offering a user-friendly interface for students and educators. By leveraging its expertise in prison education and justice reform, Nucleos offers technology-based solutions that prioritize ethical considerations and social impact. By aligning their business models with the public interest, these companies can attract socially conscious investors and customers while contributing to positive societal change through their ethical tech offerings.
- **PadSplit** is a public benefit company described as the largest coliving marketplace in the U.S. It is focused on providing affordable housing options for workers like teachers, security guards, and retail employees. Their mission is to address the affordable housing crisis while using housing as a means for financial empowerment.
- **RiseKit** is a company that aims to connect diverse and often overlooked job seekers with employers through community organizations. They offer different "kits" tailored for employers, community organizations, governments, and foundations to facilitate these connections and measure outcomes.
- **Measure** is a Black women-led data activist and community-led research non-profit organization founded in 2015. By elevating lived experiences as data and training community members as data activists, Measure builds capacity within marginalized communities to engage with and shape technology policy. They accomplish this in part by training community members and

partners in their CARE Model (Community + Advocate + Resilience + Evidence). Their work demonstrates how data and technology can be leveraged for social justice, providing a model for organizations seeking to use public interest technology to empower communities and drive policy change.

2. Public Interest Technology Partnerships

Public interest technology, as defined by the **Public Interest Technology University Network** (PIT-UN), is the application of technology expertise to advance the public good. While still emerging, particularly in the private sector, there are notable PIT success stories. In Power to the Public, Tara Dawson McGuinness and Hana Schank highlight **Civilla,** a nonprofit design studio that partnered with Michigan's Department of Health and Human Services (MDHHS) to redesign the state's benefits application form. Using human-centered design and extensive community engagement, Civilla created a more accessible application, demonstrating a model for future tech ethics initiatives.

Similarly, Amy Sample Ward and Afua Bruce in The Tech That Comes Next describe a partnership between **Rescuing Leftover Cuisine** (RLC), a nonprofit combating food waste, and **Calliope Consulting.** Calliope's long-term, community-centered approach helped RLC develop better technology systems, enabling them to scale operations significantly. These examples illustrate how technology partnerships can amplify social impact and provide models for future public interest technology endeavors.

Ward and Bruce identify a number of other partnerships between tech companies/organizations and social impact organizations that can serve as useful models for future commercialization:

- Nonprofit organizations helping mission-driven organizations leverage technology:

1. NTEN (The Nonprofit Technology Enterprise Network) is a nonprofit organization that helps other nonprofits effectively use technology to advance their missions. They provide resources, training, community support, events, webinars, conferences, courses, and certifications related to nonprofit technology.
2. DataKind is a nonprofit organization that harnesses the power of data science and AI to help mission-driven organizations tackle challenges in areas such as health, poverty, **human rights,** and the environment. They run data science projects, workshops, and events to help social good organizations use data effectively and ethically to increase their impact.

- Tech company initiatives helping mission-driven organizations leverage technology:

1. GitHub's Tech for Social Good: GitHub's Tech for Social Good initiative supports nonprofits, open source projects, and other social impact organizations by providing them with free access to GitHub's tools and resources. This includes free private repositories, team management tools, learning resources, and community support to help these organizations collaborate effectively and build software solutions to advance their missions.
2. Okta for Good: Okta's social impact initiative, Okta for Good, offers free and discounted access to Okta's identity and access management products for nonprofits, charitable organizations, and social enterprises. The program also includes community engagement where Okta employees volunteer their skills and time, and Okta makes grants and investments in nonprofits and social enterprises using technology for positive change.

3. Mission-Driven Alternatives to Big Tech

A number of technologists have left **Big Tech** companies in recent years due to growing concerns about what they viewed as unethical, discriminatory, and/or potentially dangerous ways their respective companies approached the design, development, and deployment of their products and services. Driven by a strong sense of social responsibility and a desire to put principles into practice, some opted to forge their own paths and create their own mission-driven alternatives to **Big Tech**. Whether their focus is on data privacy and security, AI **alignment** with human values, or other critical areas, they are united by a shared goal of developing technologies that prioritize user trust and **transparency**, **human rights** and the public interest, and **social justice** broadly conceived.

Some founded research institutes and nonprofit organizations that challenge the problems and power asymmetries present in **Big Tech**. A number of former Google employees, for instance, have gone on to create their own tech ethics initiatives. Tristan Harris, a former Google design ethicist, co-founded the **Center for Humane Technology**, a nonprofit organization that raises awareness about the negative impacts of social media and technology on society and promotes the development of more humane and ethical technology products. Guillaume Chaslot, a former Google engineer, founded **AlgoTransparency**, a project that aims to increase **transparency** and accountability in algorithmic decision-making. Laura Nolan, a former Google software engineer who resigned over the company's involvement in a military drone AI imaging program, has since become an advocate for the **Stop Killer Robots campaign**.

Timnit Gebru, a prominent AI ethics researcher controversially fired from Google in December 2020, later founded the **Distributed Artificial Intelligence Research Institute** (DAIR), which aims to

conduct independent, community-rooted AI research that prioritizes equity and the perspectives of marginalized groups. These may not be for-profit companies, but each plays a significant role in the growing movement to prioritize ethics in technology development.

Other technologists left **Big Tech** to design and develop technologies better aligned with ethical principles, effectively commercializing ethical alternatives to **Big Tech** products. For example, **Signal**, a secure messaging application, was created by Moxie Marlinspike, with WhatsApp co-founder Brian Acton joining after leaving Facebook over privacy concerns. Operating as a nonprofit, Signal offers end-to-end encryption and minimal data collection, contrasting sharply with Facebook's data-driven business model. The development of Signal is driven by the belief that private and secure communication is a fundamental human right, and the app's success has made it a preferred choice for journalists, activists, and privacy-conscious individuals.

Similarly, **Anthropic** was founded by researchers who left OpenAI, concerned about insufficient safeguards in advanced AI development. Established as a public benefit corporation (PBC), Anthropic employs "constitutional ethics," embedding ethical principles into AI systems from the outset. This approach contrasts with OpenAI's partnership with Microsoft, which some viewed as potentially compromising ethical standards. These examples illustrate how former Big Tech employees are commercializing ethical alternatives, prioritizing user privacy, security, and **responsible AI** development over profit-driven models.

It should be noted, however, that **Anthropic's recent partnership with Amazon** does raise concerns about the extent to which even a chartered PBC can maintain its commitment to tech ethics principles given the pressures of capital investment. Moreover,

Anthropic employs the same problematic approach as other LLM developers: it is, ultimately, Anthropic that assesses the presumed safety and **alignment** of its own models. The company's status as a PBC and its commitment to constitutional ethics still falls well short of a fully independent third-party audit that might assess the extent to which its model actually aligns with principles of justice, equity, accountability, and safety.

Nevertheless, Signal and Anthropic offer examples of how tech ethics principles and practices can be commercialized. They demonstrate that it is possible to create financially sustainable and socially responsible technologies that prioritize user privacy, security, and the **alignment** of AI with human values. By operating on a nonprofit model, Signal ensures that its primary focus remains on protecting user privacy rather than maximizing profits. Anthropic, a public benefit for-profit company, has embedded ethical considerations into its core mission and research practices. As the tech industry continues to grapple with the ethical implications of rapidly advancing technologies, Signal and Anthropic serve as examples of responsible innovation.

V. CHALLENGES AND MITIGATION STRATEGIES

A. Identifying Challenges in Commercializing Technology Ethics

There are, of course, many challenges facing the tech ethics field, especially as companies and organizations seek to commercialize ethical, public interest, and **social justice** tech products and services. These challenges are social, political, cultural, economic, technical, and regulatory in nature, underscoring the need for an ecosystem of

diverse stakeholders engaged in a comprehensive and collaborative approach to address them.

As discussed at length above, there is a pressing need for robust regulatory frameworks to hold private companies accountable and enforce adherence to technology ethics principles and practices. The current landscape is characterized by a lack of standardized regulations and guidelines, which creates uncertainty and hinders the adoption of tech ethics. Even where regulations do exist, enforcing compliance and ensuring ethical practices can be challenging, particularly for **small and medium businesses** and those operating across multiple jurisdictions. The fact that there will continue to be multiple competing regulatory frameworks operating on a global scale and impacting businesses around the world makes this all the more challenging. Without a strong regulatory system, companies may prioritize short-term profits over long-term ethical considerations, perpetuating potential harm and undermining public trust.

Closely tied to regulatory challenges is the need to establish professional industry-wide standards and independent, third-party certification processes for tech ethics practitioners. Organizations like the IEEE and the IAAA have begun this work, but until there are widely accepted auditing frameworks with clear independent oversight for auditors, there will continue to be concerns about consistency, credibility, and the potential for **"ethics washing."** Without established enforcement mechanisms, companies may engage in superficial audits solely for public relations purposes, without a genuine commitment to addressing ethical concerns. Developing industry-wide standards, certification processes, and regulatory frameworks for **algorithmic auditing** and tech ethics services is crucial to ensure their integrity and effectiveness.

As highlighted in the Shadow Report, there is a concerning trend of policymakers and industry leaders prioritizing corporate interests over public concerns in AI governance discussions. Despite strong public support for AI regulation and growing concerns about AI's societal impacts, many forums and policy discussions have centered corporate priorities rather than public interest. This **elite capture** can further contribute to superficial or performative adoption of ethical language or practices by organizations to appear socially responsible without genuine commitment. The Shadow Report warns that even formally independent assessors and auditors can become dependent on maintaining favorable relationships with industry, potentially softening their evaluations.

The field faces other social, cultural, and political challenges as well. As important ideas and approaches in the public interest and **social justice** tech space (like "critical race theory," "diversity, equity, and inclusion," and even the term "**social justice**" itself) are captured and weaponized in the context of cultural and political polarization, this will make it all the more challenging to demonstrate the importance of these products and services in certain quarters. Not only that, building public understanding and trust will require concerted efforts to engage diverse stakeholders, particularly minoritized and vulnerable communities that have been disproportionately impacted by unethical tech practices. There are long histories of broken trust between tech companies and vulnerable communities, broken trust rooted in longer histories of settler colonialism, white supremacy, misogyny, and other histories of violence and institutions of oppression. There is, thus, a sociocultural imperative to not only to aggregate demand and raise public consciousness around tech ethics and public interest technology, but also to ensure that this work is spearheaded by members of those very marginalized communities most impacted by these technologies and their social impacts.

Additional political and economic challenges that shape the tech ethics sector include:

1. Varying priorities and political interests: The prioritization of tech ethics and the allocation of resources to support these initiatives can be influenced by changing political landscapes and varying priorities of different administrations or governments.
2. Potential resistance from stakeholders: Some corporate stakeholders or government agencies may resist the adoption of tech ethics services and products if they perceive them as hindering their operations or imposing additional costs and burdens.
3. Cost and resource constraints: Many of the proposed tech ethics services and products, such as **algorithmic auditing**, third-party **vetting**, and workforce development, require significant resources and investment. Smaller organizations or startups may struggle to afford these services.

B. Mitigation Strategies for Overcoming Hurdles

Overcoming these challenges demands a collaborative approach involving policymakers, industry leaders, academics, activists and advocates, civic organizations, and the public at large. This collaborative ecosystem should prioritize not just ethics but even more so **social justice** and the public interest. It should work to establish clear standards and regulatory frameworks while promoting **transparency** and public accountability. By doing so, the tech ethics field can navigate the complexities of the current landscape and realize its full potential in shaping the responsible development and deployment of emerging technologies.

Potential mitigation strategies include:

1. Creative artistic, activist, and advocacy efforts to build public pressure for regulation

- Collaborations with artists, writers, and performers to create compelling narratives and artistic expressions that highlight the importance of tech ethics and the potential consequences of unregulated technology.
- Support for advocacy campaigns and grassroots movements that demand accountability and ethical practices from tech companies and policymakers.
- Leveraging the power of storytelling and emotional connections to raise public awareness and mobilize broader support for tech ethics regulation.

2. Support and collaboration with professional associations to standardize tech ethics:

- Working closely with organizations like the IEEE, the IAAA, and other professional bodies to develop industry-wide standards and best practices for tech ethics services, such as **algorithmic auditing.**
- Contributing to the creation of certification processes and training programs to ensure the competence and credibility of tech ethics professionals.
- Advocating for the adoption of these standards and certifications by companies, regulatory bodies, and policymakers, fostering accountability and consistency.

3. Advocate for the creation of a “Tech Ethics Maturity Model” for government contractors:

- The Department of Defense's Cybersecurity Maturity Model Certification (CMMC) offers a potential model for addressing the lack of enforceable standards and industry-wide benchmarks for tech ethical practices.
- The CMMC requires defense contractors to meet specific cybersecurity standards, with different levels of certification based on the sensitivity of information handled. This tiered approach, coupled with third-party audits, ensures a baseline of security practices across the defense industrial base.
- An analogous “Tech Ethics Maturity Model” might require companies to demonstrate increasing levels of commitment to fairness, **transparency,** and community benefit in their tech solutions.
  Level 1 might require basic ethical guidelines and employee training
  Level 2 might also require regular internal audits and public reporting
  Levels 3 and **above** might require third-party ethical audits, advanced bias mitigation strategies, proven track records of positive social impact, etc.
- This approach could leverage government procurement power to create market incentives for companies to prioritize ethical considerations, standardize best practices, and provide a framework for assessing a company's ethical maturity.

4. Building an ecosystem of diverse stakeholders to raise awareness and aggregate demand:

- Fostering public-private partnerships and collaborations with civic organizations, academia, and policymakers.
- Emphasizing the interdisciplinary nature of tech ethics by actively engaging stakeholders not only from STEM fields, but also from the social sciences, humanities, arts, and other relevant fields.
- Promoting a humanities-driven approach to tech ethics that recognizes the value of cross-cultural empathy, critical thinking, epistemic humility, and nuanced perspectives offered by the humanities.
- Adopting a public interest technology (PIT) approach that prioritizes the design, data, and delivery of technologies for the greater good of society, particularly underserved and marginalized communities.
- Embracing an **ecosystem approach** to **sociotechnical** problem-solving, acknowledging the interconnectedness of technological, social, cultural, and ethical considerations.

By implementing these strategies, the tech ethics field can overcome the challenges it faces and create a supportive ecosystem for the responsible development and deployment of emerging technologies. Building public pressure, standardizing practices, and fostering multidisciplinary collaborations can drive regulatory changes, establish professional credibility, and raise awareness about the importance of tech ethics in shaping a more equitable and ethical technological future.

VI. FUTURE TRENDS AND INNOVATIONS

Anticipating future trends in technology innovation and responsible **governance** requires an anticipatory understanding of emerging developments and their ethical implications. Several pivotal areas are poised to significantly influence the trajectory of technology innovation in the coming years:

Spatial Computing

This encompasses various immersive technologies such as virtual reality (VR), augmented reality (AR), and AI-driven experiences. As technological capabilities rapidly evolve, spatial computing will permeate diverse sectors such as entertainment, education, healthcare, and architecture. For instance, advancements in AR could revolutionize remote training programs for professionals in industries like engineering and aviation. It will also heighten and potentially exacerbate the patterns of race- and gender-based violence in gaming environments.

Agricultural Technology

With the imperative to ensure global food security, agricultural technology assumes a critical role. Innovations like precision farming, vertical farming, and automated crop monitoring systems will be pivotal in addressing food production challenges. Additionally, the development of plant-based meat alternatives and advanced applications of genetically modified food crops represent a burgeoning field with significant ethical considerations, particularly concerning sustainability, food security, and animal welfare.

Financial Technology (Fintech)

The fintech landscape is undergoing rapid transformation, propelled by innovations such as digital currencies, blockchain technology, and algorithmic trading. These advancements are reshaping traditional financial systems and democratizing access to banking services. Moreover, decentralized finance (DeFi) platforms are emerging as disruptive alternatives to traditional banking, enabling peer-to-peer lending and transparent financial transactions. Both within and across national boundaries, the challenges for equity, accountability, and vetted protocols will be significant and will create significant opportunities for PIT-related services.

Genetic Engineering

Breakthroughs in genetic engineering hold immense promise for revolutionizing healthcare and agriculture. Technologies such as CRISPR-Cas9 gene editing offer unprecedented precision in modifying genetic material, leading to potential cures for genetic disorders and advancements in crop breeding. The FDA approved a CRISPR-based cure for sickle-cell anemia in 2023, and additional therapies are likely to follow soon. However, ethical considerations surrounding gene editing in humans, including concerns about unintended consequences and equitable access to genetic therapies, remain paramount and may easily outstrip the disruption currently witnessed in the AI sector. The production of bio-weapons using genetic engineering and AI, moreover, will easily destabilize geopolitics and potentially devastate highly vulnerable **populations** without robust **governance** and ethical controls in place.

Environmental Technology

As the consequences of climate change become increasingly dire, technological solutions play a pivotal role in mitigating environmental risks. Innovations such as renewable energy systems

and sustainable waste management practices are essential for fostering a greener future. Additionally, advancements in biodegradable materials and circular economy initiatives present opportunities for reducing environmental degradation and promoting ecological resilience. The policies and regulatory frameworks for these issues will be urgent.

Space Exploration and Defense

The burgeoning commercial space industry, coupled with geopolitical tensions, underscores the need for ethical **governance** in space activities. From satellite constellations for global internet connectivity to lunar mining ventures, commercial space ventures present both opportunities and challenges that may reproduce the most disturbing histories of earth-based colonialism, apartheid, and extraction. Moreover, the development of space-based technologies for Earth observation and disaster management underscores the ethical imperative of ensuring equitable access to space resources and mitigating orbital debris pollution.

Transnational Governance Systems

In an increasingly interconnected world, effective **governance** mechanisms are essential for regulating emerging technologies on a global scale. International bodies such as the United Nations play a pivotal role in setting standards and norms for responsible technology use. Moreover, initiatives like the Global Partnership on Artificial Intelligence (GPAI) aim to foster international cooperation in addressing ethical challenges posed by AI and other disruptive technologies. Creating more robust systems of global technology **governance** will become a critical area for PIT-related services.

VII. CONCLUSION

A. Summary of Key Findings

This white paper has provided an overview of the current state of the **technology ethics** field, highlighting the growing importance of ethical considerations in the development and deployment of AI and other emerging technologies. It has identified a wide range of commercialization opportunities across the tech lifecycle, from **algorithmic auditing** and **data ethics** consulting to workforce development and public interest technology partnerships as well as a host of tech ethics needs that will continue to emerge in specific sectors of the economy. The case studies and success stories cited showcase the viability and impact of mission-driven tech nonprofits, companies, and public benefit corporations in advancing innovation. However, there are also many multifaceted challenges facing the field, including the need for robust regulatory frameworks, industry-wide standards, and continued efforts to raise public awareness and build public trust. As stakeholders navigate this complex landscape, perhaps above all, it will be crucial for them to clearly identify whether they are approaching tech ethics work from the vantage point of a more general conception of "responsible tech" ethics, or whether they are explicitly emphasizing a public interest and **social justice** tech approach. This distinction will shape their priorities, strategies, and the ultimate impact of their efforts in building a more equitable and ethical technological future.

B. Encouraging Ethical Practices as a Business Imperative

As the industry continues to shape our world in profound ways, prioritizing ethical practices is no longer a mere option — it is a business imperative.

Companies that fail to address the societal implications of their products and services risk severe reputational damage, loss of public trust, and regulatory penalties. In an era of heightened scrutiny and growing public awareness, businesses that neglect their ethical responsibilities may face consumer boycotts, employee protests, and legal liabilities. Moreover, as regulatory frameworks such as the **EU AI Act** and the White House AI Executive Order come into effect, non-compliance with ethical standards could result in substantial fines, sanctions, and restrictions on market access.

Conversely, companies that proactively embed ethical principles into their **governance**, design, and deployment processes stand to gain a competitive advantage by building consumer confidence, attracting top talent, and contributing to positive societal impact. By prioritizing values such as **transparency,** accountability, and fairness, businesses can foster trust among their stakeholders and differentiate themselves in an increasingly crowded market. Ethical practices can also drive innovation by encouraging the development of socially responsible and sustainable technologies that address pressing global challenges. Furthermore, by attracting and retaining employees who are passionate about making a positive difference, companies can cultivate a purpose-driven culture that enhances motivation, productivity, and loyalty.

The commercialization of **technology ethics** services and products presents a significant opportunity for businesses to align their financial interests with the greater good. By offering services such as **algorithmic auditing**, **data ethics** consulting, and public interest technology partnerships, companies can tap into growing market demand while contributing to the development of a more ethical and equitable technology ecosystem. This approach not only mitigates risks but also unlocks new revenue streams and enhances

brand reputation. As consumers become increasingly conscious of the ethical implications of their purchasing decisions, businesses that demonstrate a genuine commitment to responsible innovation will be well-positioned to thrive in the long term.

C. Call to Action for Stakeholders in the Technology Industry

Now is arguably the most urgent time for decisive action on the part of stakeholders committed to public interest technology. The opportunities offered by advancements in technology are tremendous. As this study indicates, the potential benefits of technology innovation are rivaled by the actual and potential harms associated with these same technologies. These developments have been met, fortunately, by an unprecedented scale of awareness and strategy devoted to foregrounding public wellbeing and making technology accountable to the best interests of society.

The complex conditions that shape ethical challenges and societal concerns for technology constitute a vast opportunity for public interest technology to be more translational by providing products and services through a sustainable, ethically guided business model. Governments throughout the globe are now familiar with the importance of regulating technology. **Human rights** organizations uniformly recognize that safeguarding civil liberties and fundamental rights requires engagement with the rapidly evolving innovation landscape. And private companies have begun procuring services to support their ability to comply with best practices and the emerging policies and guidelines that will be essential to technology **governance.**

The coming years will likely prove decisive for determining how technology can be made more accountable to public interest and

societal benefit by expanding public interest technology more fully into the commercial space. Through various forms of consulting and provisioning products and services that foster more equitable outcomes, public interest technology can embrace a pivotal opportunity. In the future, ethically aligned commercialism must play a more central role in guiding our technological world. This report is meant to elevate the likelihood of success for that immense endeavor.

VIII. APPENDICES

A. Glossary of Terms

- **AI EO (White House Executive Order on AI):** The White House Executive Order on the Safe, Secure, and Trustworthy Development and Use of AI issued by President Joseph R. Biden in October 2023.
- **Algorithm:** A set of step-by-step instructions or rules that a computer system follows to perform a specific task, solve a problem, or make a decision, often forming the basis for and machine learning processes in AI systems.
- **Algorithmic auditing:** The process of examining an algorithmic system for bias, fairness, transparency, and adherence to ethical standards.
- **Algorithmic bias:** Systematic errors in algorithmic systems that lead to unfair or discriminatory outcomes, often due to biased training data or model design.
- **Alignment:** Ensuring that AI systems are designed and operate in ways that are consistent with human values, goals, and ethical principles.
- **Artificial intelligence (AI):** The development of computer systems that can perform tasks that typically require human intelligence.
- **Automated decision-making:** The use of algorithms to make decisions without direct human involvement.
- **Big Tech:** The five largest, most dominant, and influential technology companies—Alphabet (Google), Amazon, Apple, Meta (Facebook), and Microsoft—which collectively wield significant economic, social, and political power.

- **Biometric data:** Unique physical or behavioral characteristics used for identification, such as fingerprints, facial features, or voice patterns.
- **Black box:** A term used to describe AI systems or algorithms that make decisions or produce outputs without providing clear explanations of their internal workings, making it difficult for users, developers, and regulators to understand, audit, or challenge their functioning and potential biases.
- **Constitutional AI:** AI systems designed with built-in ethical principles and constraints from the ground up.
- **Corrigibility:** The ability of an AI system to be corrected, interrupted, or shut down by humans if it behaves in undesirable or harmful ways.
- **Cyberspace Administration of China:** The central government agency responsible for overseeing and regulating China's internet and digital industries, including the development and implementation of policies related to online content, data security, and emerging technologies such as artificial intelligence.
- **Data ethics:** The study and application of ethical principles to the collection, analysis, and use of data.
- **Deepfakes:** AI-generated synthetic media, such as images, videos, or audio, that appear authentic but are fabricated or manipulated.
- **Digital discrimination:** Unfair treatment or bias based on personal characteristics in digital environments or through the use of digital technologies.
- **Dual-use dilemma:** Technologies that can be used for both beneficial and harmful purposes.
- **Ecosystem approach:** An approach to the problem space that calls for diverse stakeholders to collaborate in addressing the social, ethical, legal, and economic implications of technological development and deployment.

- **Elite capture:** A process by which powerful, privileged, or influential individuals or groups seize control over decision-making processes, resources, or narratives, often by leveraging their existing advantages or exploiting systemic inequalities, ultimately redirecting benefits and outcomes to serve their own interests at the expense of the broader population or marginalized communities.
- **Ethics:** The study of moral principles and values that govern human behavior and decision-making.
- **Ethics washing:** Superficial or performative adoption of ethical language or practices by organizations to appear socially responsible without genuine commitment.
- **EU AI Act:** The European Union regulation adopted on March 13, 2024 that aims to create a comprehensive legal framework for AI, focusing on risk management and compliance.
- **Explainable AI:** AI systems designed to provide clear, understandable explanations of their decision-making processes and outputs.
- **Facial recognition:** Technology that can identify or verify individuals based on their facial features.
- Foundation models: Large AI models, often trained on vast amounts of data, that can be adapted for various downstream tasks.
- **Generative AI (GenAI):** AI systems that can create new content, such as text, images, or music, based on patterns learned from training data.
- **Ghost work:** Invisible or unrecognized human labor that supports AI systems, such as data labeling or content moderation.
- **Governance:** The processes, structures, and practices used to oversee, manage, and control the development and use of technology.

- **High-risk AI systems:** AI applications that pose significant risks to human rights, health, safety, or fundamental freedoms, as defined by the EU AI Act.
- **Human rights:** Fundamental rights and freedoms that every person is entitled to, regardless of their background or characteristics.
- **Human-centered AI:** AI systems designed with human needs, values, and well-being as the central focus.
- **Interpretability:** The ability to explain and understand how an AI system works and makes decisions.
- **Large language models (LLMs):** AI models trained on vast amounts of text data that can generate human-like text and perform various language tasks.
- **Machine learning:** A subfield of AI that involves training computer systems to learn and improve from data without being explicitly programmed.
- **Multimodal AI:** AI systems that can process and integrate multiple types of data, such as text, images, and audio, to perform tasks or make decisions.
- **Predictive policing:** The use of AI and data analytics to predict and prevent potential crimes or identify likely offenders.
- **Public benefit corporations (PBCs):** Corporate entities that consider the interests of multiple stakeholders and commit to creating positive social and environmental impact.
- **Public interest technology (PIT):** The development and use of technology to promote the public good and address societal challenges.
- **Responsible AI:** The practice of developing and using AI systems in an ethical, transparent, and accountable manner.
- **Risk assessment:** The process of identifying, evaluating, and mitigating potential risks associated with the development and use of AI systems.

- **Robustness:** The ability of an AI system to maintain performance and reliability under varying conditions or when facing unexpected inputs.
- **Scalable oversight:** Techniques for monitoring and controlling AI systems that can be applied efficiently to large-scale deployments.
- **Small and medium businesses (SMBs):** Companies with limited resources and smaller-scale operations compared to large corporations.
- **Social justice:** The fair and equitable distribution of wealth, access to resources and opportunities, and protection of rights and privileges within a society.
- **Social justice technology:** Technologies designed and used to promote social justice, equity, and inclusion.
- **Sociotechnical:** An approach that recognizes the complex and interdependent relationships between technology and society, emphasizing the importance of considering social, cultural, political, and ethical factors alongside technical aspects when designing, developing, and deploying technological systems.
- **Surveillance capitalism:** A business model that relies on the collection, analysis, and monetization of personal data for profit.
- **Synthetic data:** Artificially generated data that mimics real-world data patterns and characteristics.
- **Technology ethics:** The study and application of moral principles to the development, use, and impact of technology on society.
- **Technological solutionism:** The belief that technology can be the primary or sole solution to complex social, political, or economic problems, often leading to an overemphasis on technical interventions while neglecting the importance of addressing underlying systemic issues and potential unintended consequences.

- **Third-party evaluation/auditing:** Independent assessment of AI systems by external experts to ensure compliance with ethical standards and regulations.
- **Transparency:** The practice of openly communicating information about the design, development, and use of AI systems.
- **Trustworthy AI:** AI systems that are reliable, safe, transparent, and aligned with human values and ethical principles.
- **Vetting:** The process of thoroughly evaluating and verifying the safety, security, and ethical compliance of AI systems before deployment.

B. Works Cited (In Order Of Citation)

- Kate Crawford, Atlas of *AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* (New Haven: Yale University Press, 2021).
- Tara Dawson McGuinness and Hana Schank, *Power to the Public: The Promise of Public Interest Technology* (Princeton: Princeton University Press, 2021).
- Digital Regulation Cooperation Forum, "Auditing Algorithms: The Existing Landscape & Future Developments in the Algorithmic Auditing Space," September 2022, https://www.gov.uk/government/publications/findings-from-the-drcf-algorithmic-processing-workstream-spring-2022/auditing-algorithms-the-existing-landscape-role-of-regulators-and-future-outlook.
- Amy Sample Ward and Afua Bruce, *The Tech That Comes Next: How Changemakers, Philanthropists, and Technologists Can Build an Equitable World* (Hoboken: Wiley, 2022).

C. Additional Resources For Further Reading

- Amrute, Sareeta. Encoding Race, Encoding Class: *Indian IT Workers in Berlin.* Duke University Press: 2016.
- Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code.* Cambridge, UK: Polity, 2019.
- Birhane, Abeba, Pratyusha Kalluri, Dallas Card, William Agnew, Ravit Dotan, and Michelle Bao. "The Values Encoded in Machine Learning Research." arXiv:2106.15590 (2021).
- Broussard, Meredith. *Artificial Unintelligence: How Computers Misunderstand the World.* Cambridge, MA: MIT Press, 2018.
- Buolamwini, *Joy. Unmasking AI: My Mission to Protect What is Human in a World of Machines.* New York: Penguin Random House, 2023.
- Costanza-Chock, Sasha. Design *Justice: Community-Led Practices to Build the Worlds We Need.* Cambridge, MA: MIT Press, 2020.
- Donig, Deb. The Technically Human Podcast.
- Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor.* New York: St. Martin's Press, 2018.
- Kearns, Michael, and Aaron Roth. The Ethical Algorithm: *The Science of Socially Aware Algorithm Design.* New York: Oxford University Press, 2019.
- Lamdan, Sarah. Data Cartels: *The Companies That Control and Monopolize Our Information.* Stanford, CA: Stanford University Press, 2022.
- Mohamed, Shakir, Marie-Therese Png, and William Isaac. "Decolonial AI: Decolonial Theory as Sociotechnical Foresight in Artificial Intelligence." *Philosophy & Technology* 33, no. 4 (2020): 659-684.

- Noble, Safiya Umoja. *Algorithms of Oppression*: *How Search Engines Reinforce Racism.* New York: NYU Press, 2018.
- O'Neil, Cathy. Weapons of Math Destruction: *How Big Data Increases Inequality and Threatens Democracy.* New York: Crown, 2016.
- Pasquale, Frank. The Black Box Society: *The Secret Algorithms That Control Money and Information.* Cambridge, MA: Harvard University Press, 2015.
- Schneider, Nathan. Governable Spaces: Democratic Design for Online Life. Berkeley, CA: University of California Press, 2024.
- Vallor, Shannon. Technology and the Virtues: *A Philosophical Guide to a Future Worth Wanting.* New York: Oxford University Press, 2016.
- Viljoen, Salomé. "A Relational Theory of Data Governance." Yale Law Journal 131, no. 2 (2021): 573-654.
- Wachter-Boettcher, Sara. *Technically Wrong*: *Sexist Apps, Biased Algorithms, and Other Threats of Toxic Tech.* New York: W. W. Norton & Company, 2017.
- Zuboff, Shoshana. The Age of Surveillance Capitalism: *The Fight for a Human Future at the New Frontier of Power.* New York: Public Affairs, 2019.

“The viability of technology, like democracy, depends in the end on the practice of justice and on the enforcement of limits to power.” –Ursual Franklin

ACKNOWLEDGEMENTS

The insights of this landscape survey have benefited greatly from conversations with key stakeholders in the social justice tech ecosystem. We are deeply grateful for their willingness to speak with us and contribute to our understanding of the field. The photos and elements in this version were generated using the Canva platform solely for presentation purposes.

Prepared by the
Corporation for Public Interest Technology
for the New Venture Fund and the
Public Interest Technology-University Network

CONTACT

Corporation for Public Interest Technology

4315 50th Street NW, Suite 100 #7265, Washington, DC 20016

www.cpitpbc.com | contact@cpitpbc.com