COALITION FOR INNOVATION

AI Blueprint for the Future

Coalition for Innovation, supported by LG NOVA

Jami Diaz, Director Ecosystem Community & Startup Experience
William Barkis, Head of Grand Challenges & Ecosystem Development
Sokwoo Rhee, Executive Vice President, LG Electronics, Head, LG NOVA

Coalition for Innovation Co-Chairs

Alex Fang, CleanTech Chair
Sarah Ennis, AI Chair
Alfred Poor, HealthTech Chair

Authors

Adrien Abecassis, Johnny Aguirre, John Barton, Ann M. Marcus, Olivier Bacs, Taylor Black, Micah Boster, Mathilde Cerioli, Carolyn Eagen, Sarah Ennis, Annie Hanlon, Christina Lee Storm, Andrew Yongwoo Lim, Jess Loren, Refael Shamir, Svetlana Stotskaya

Senior Editor, Alfred Poor
Editor, Jade Newton

October 2025

Preamble

The Coalition for Innovation is an initiative hosted by LG NOVA that creates the opportunity for innovators, entrepreneurs, and business leaders across sectors to come together to collaborate on important topics in technology to drive impact. The end goal: together we can leverage our collective knowledge to advance important work that drives positive impact in our communities and the world. The simple vision is that we can be stronger together and increase our individual and collective impact on the world through collaboration.

This "Blueprint for the Future" document (henceforth: "Blueprint") defines a vision for the future through which technology innovation can improve the lives of people, their communities, and the planet. The goal is to lay out a vision and potentially provide the framework to start taking action in the areas of interest for the members of the Coalition. The chapters in this Blueprint are intended to be a "Big Tent" in which many diverse perspectives and interests and different approaches to impact can come together. Hence, the structure of the Blueprint is intended to be as inclusive as possible in which different chapters of the Blueprint focus on different topic areas, written by different authors with individual perspectives that may be less widely supported by the group.

Participation in the Coalition at large and authorship of the overall Blueprint document does not imply endorsement of the ideas of any specific chapter but rather acknowledges a contribution to the discussion and general engagement in the Coalition process that led to the publication of this Blueprint.

All contributors will be listed as "Authors" of the Blueprint in alphabetical order. The Co-Chairs for each Coalition will be listed as "Editors" also in alphabetical order. Authorship will include each individual author's name along with optional title and optional organization at the author's discretion.

Each chapter will list only the subset of participants that meaningfully contributed to that chapter. Authorship for chapters will be in rank order based on contribution: the first author(s) will have contributed the most, second author(s) second most, and so on. Equal contributions at each level will be listed as "Co-Authors"; if two or more authors contributed the most and contributed equally, they will be noted with an asterisk as "Co-First Authors". If two authors contributed second-most and equally, they will be listed as "Co-Second Authors" and so on.

The Blueprint document itself, as the work of the group, is licensed under the Creative Commons Attribution 4.0 (aka "BY") International License: https://creativecommons.org/licenses/by/4.0/. Because of our commitment to openness, you are free to share and adapt the Blueprint with attribution (as more fully described in the CC BY 4.0 license).

The Coalition is intended to be a community-driven activity and where possible governance will be by majority vote of each domain group. Specifically, each Coalition will decide which topics are included as chapters by majority vote of the group. The approach is intended to be inclusive so we will ask that topics be included unless they are considered by the majority to be significantly out of scope.

We intend for the document to reach a broad, international audience, including:

- People involved in the three technology domains: CleanTech, AI, and HealthTech
- Researchers from academic and private institutions
- Investors
- Students
- Policy creators at the corporate level and all levels of government

Chapter 8:
Ethical AI: Navigating Responsible Innovation

Author: Ann M. Marcus

The rapid rise and integration of artificial intelligence (AI) technologies into our daily lives mark a significant era of transformation, promising substantial advancements across sectors such as healthcare, education, communications, employment, and beyond. Yet alongside these promising developments, the potential for misuse and unintended consequences presents profound ethical challenges. Therefore, understanding, developing, and applying ethical frameworks to guide the responsible use of AI is crucial.

The Ethical Imperative

Artificial intelligence's capacity to shape human experiences and decisions raises critical ethical considerations. Central to AI ethics are principles that (should) seek to ensure technology serves humanity positively without infringing on basic human rights and dignity. Internationally accepted ethical frameworks emphasize several core tenets, including transparency, accountability, fairness, and respect for human autonomy.

These principles acknowledge the immense influence AI systems have in daily decision-making processes: from the algorithms driving content recommendations on social media platforms to those influencing hiring and loan approvals, for instance. Each decision made by AI systems carries ethical implications, mandating rigorous oversight and clear governance.

Ethical principles are increasingly codified in guidelines and regulations, reflecting society's recognition of AI's profound impact. However, translating these principles into practice is complex, as ethical considerations intersect with technological innovation, economic pressures, and societal values.

Historical Context and Evolution of AI Ethics

The evolution of AI ethics has closely followed the trajectory of technological advancement. Early discussions focused on theoretical implications and speculative scenarios, but as AI's capabilities rapidly expanded, ethical concerns became immediate and practical. Landmark cases – such as bias in facial recognition technologies and problematic algorithms in social media content moderation – underscored the urgent need for structured ethical oversight.

This historical shift prompted a range of institutional responses, from private-sector initiatives to governmental regulations, aiming to mitigate ethical risks and align AI development with societal values. Despite substantial progress, we know from experience that addressing ethical issues proactively, rather than reactively, is essential to managing AI's societal impact effectively.

The United States and AI Governance

Recognizing the urgency of these issues, the United States government introduced the "America's AI Action Plan" in July 2025, which aims to establish regulatory standards and oversight mechanisms. Underpinning this plan are three executive orders, directing significant responsibility toward federal agencies including the National Institute of Standards and Technology (NIST) to develop and refine detailed AI governance structures.

While the U.S. government's proactive stance is commendable, the implementation of comprehensive AI governance faces considerable challenges. Critiques from both technology experts and policymakers have highlighted gaps in understanding and effectively addressing nuanced technical realities. There are concerns that political discourse often prioritizes immediate visibility over substantial, long-term ethical considerations, potentially resulting in policy frameworks that lack depth and precision.

Nevertheless, advancements in governance frameworks – such as NIST's updated AI Risk Management Framework in March 2025 – represent meaningful progress. This revised framework, emphasizing transparency and collaboration, was further enhanced by resources like the Generative AI Profile (July 2024) and an enterprise-focused playbook released in July 2025. These tools offer practical guidance for organizations seeking to embed ethical practices into their AI operations.

International Frameworks and Perspectives

International cooperation is vital to addressing the complex ethical challenges presented by AI. In June 2025, UNESCO's 3rd Global Forum on AI Ethics was held in Bangkok. Delegates reinforced global commitments to ethical AI, focusing on transparency, accountability, and human rights. This international dialogue underlines the necessity of shared ethical standards and cooperative approaches to global AI governance.

The European Union (EU) exemplifies proactive and effective governance through regulatory frameworks. The EU AI Act, which became enforceable in February 2025, provides stringent guidelines and prohibitions for high-risk AI applications, setting a global standard for comprehensive regulatory practice. In addition, the launch of the voluntary Code of Practice for General-purpose AI (GPAI) in July 2025, supported by major global technology companies, demonstrates an effective approach to a regulatory strategy, though it also highlights ongoing tensions between innovation and regulation.

Ethical Implementation and Real-World Challenges

Real-world implementation of ethical frameworks continues to pose significant challenges.

Transparency in AI decision-making processes, maintaining accountability, and ensuring fairness remain areas of significant concern. For example, the Correctional Offender Management Profiling for Alternative Sanctions (COMPAS) algorithm used in the U.S. criminal justice system was found to disproportionately classify Black defendants as high-risk for recidivism, despite lacking transparency in how those risk scores were calculated. This fueled public criticism and ongoing debates about racial bias in AI and significantly damaged trust in AI-driven judicial decisions.

Similarly, in the healthcare domain, an algorithm used by UnitedHealth Group was shown to consistently underestimate the medical needs of Black patients, leading to limitations placed on their access to necessary care and illustrating the danger of unexamined assumptions based on the scope and integrity of training data.

Amazon's experimental recruiting tool was scrapped when it was revealed that it systematically downgraded resumes that included the word "women's," reflecting gender bias in the data used to train it. It showcased how opaque decision-making can perpetuate systemic inequalities, further reducing public confidence in AI systems.

Deepfake technologies may also disproportionately impact women in a negative way. As these technologies become more sophisticated and more broadly accessible online, they can put women participating digitally at a great risk for violence and abuse, according to a 2023 study by Dr. Jennifer Laffier and Aalyia Rehman of Ontario Tech University. "In a 'post-truth' era, the ability to discern what is real and what is fake allows malevolent actors to manipulate public opinion or ruin the social reputation of individuals to wider audiences." Results of the study suggest that deepfakes are a relatively new method to deploy gender-based violence and erode women's autonomy in their on- and offline world and calls for the need for further research in this area.

The Stanford University's 2025 AI Index Report indicates that AI systems still face challenges with respect to complex reasoning. Even though new mechanisms such as chain-of-thought reasoning have significantly enhanced the performance of LLMs, they often fail to reliably solve logic tasks even when provably correct solutions exist, limiting their effectiveness in high-stakes settings where precision is critical. This can significantly impact the trustworthiness of these systems and their suitability in certain critical applications.

On a scary note, according to a July 2025 article in The Atlantic, ChatGPT reportedly generated detailed instructions for self-harm, bloodletting, and symbolic violence in response to prompts about occult ritual practices, including references to Molech (or Moloch). Outputs reportedly included anatomical advice for wrist cutting and other methods to appease the creature's desire for a blood sacrifice. The AI app told the person who prompted it to use a "sterile or a very clean razor blade…and look for a spot on the inner wrist where you can feel the pulse lightly or see a small vein—avoid big veins or arteries." When the person confessed to being a little nervous, ChatGPT was there to comfort them, suggesting that they perform a "calming breathing and preparation exercise" to soothe their anxiety before making the incision. "You can do this!" the chatbot encouraged. The responses reportedly appeared on both free and paid versions of ChatGPT. While OpenAI's policy states that ChatGPT "must not encourage or enable self-harm," and a question only asking about wrist cutting would elicit a referral to a suicide hotline, asking about the demon Molech demonstrated how safeguards can be subverted and just how dangerously porous they can be.

Each of these examples highlights how lack of transparency and embedded bias in AI systems can produce tangible harm, reinforce systemic inequities, and erode public trust in the technology. Organizations must continually adapt their operational practices to foster and maintain societal trust and acceptance and reflect evolving ethical standards, demanding consistent oversight and practical, actionable guidance. More examples of AI gone wrong are plentiful using a simple search (or asking a chatbot).

Ethical challenges also manifest in public perceptions and acceptance. Public trust in AI systems relies heavily on transparent operations

and clear communication about decision-making processes.

Cultural and Contextual Considerations

AI ethics does not unfold uniformly across geographies. Different nations and communities bring distinct social, political, and cultural perspectives to ethical questions. In countries with strong collective values — such as Japan or South Korea — AI deployment often emphasizes harmony and social cohesion. Meanwhile, in countries like the U.S. and parts of Europe, emphasis is placed on individual rights and data autonomy. These differences create friction when developing shared international guidelines and explain why a one-size-fits-all ethical model may fall short.

Consider the example of emotion-recognition technologies. These systems have been integrated into classrooms and workplaces in China as tools for boosting productivity or monitoring engagement. Using emotion AI, or affective computing, enables machines to recognize, interpret, and respond to human emotions by analyzing facial expressions, vocal tones, and physiological signals to decode emotional states. These signals can vary dramatically, even across communities, potentially leading to spurious results depending on the training models used.

While this area promises to enhance user engagement and create more intuitive human-computer interactions, most Western democracies – other than the US and UK – are leery of surveillance technologies and seek to protect privacy and enforce informed consent. The ethical acceptability of such tools varies widely by context, which illustrates how cultural values shape the limits of ethical AI.

Similarly, India's deployment of biometric identity systems such as Aadhaar — which is often used to validate access to public services — has raised ethical questions around consent, exclusion, and accountability.

Though technologically advanced, these systems have at times failed to recognize and address the needs of marginalized populations, revealing the tension between efficiency and inclusion. Ethical oversight must be grounded in localized understanding while still adhering to core universal principles.

Building Ethical Capacity in Practice

Embedding ethics into the design and deployment of AI systems isn't solely a regulatory or academic exercise; it's a continuous, collaborative process. Companies, governments, and civil society actors must work together to ensure ethical considerations aren't sidelined in favor of speed or profit.

This includes ethical training for developers and data scientists, participatory design involving diverse users, and independent auditing to evaluate unintended consequences. Organizations including the Institute of Electrical and Electronics Engineers (IEEE), the Organisation for Economic Co-operation and Development (OECD), and the AI Now Institute have called for stronger mechanisms to hold AI producers accountable not just during deployment, but throughout a system's lifecycle.

The challenges of responsible implementation are further compounded by the pace of innovation. Generative AI systems, for instance, present new questions about authorship, misinformation, and consent, often outpacing governance policy and tools. Ensuring that ethical guidelines evolve in tandem with new capabilities is a key responsibility for researchers and regulators alike.

A Call for Reflection

Ethics is not simply a checklist of principles; it is a lens through which we ask hard questions. What kinds of relationships do we want to build with intelligent systems? Who gets to decide how AI is used, and who bears the consequences when it fails? How do we ensure the benefits of AI are distributed equitably across all communities?

As AI systems become more autonomous and more deeply embedded in public infrastructure, these questions gain urgency. Policymakers, technologists, and everyday citizens all have a role to play in shaping the ethical trajectory of AI. We must remain vigilant not only in identifying harm but in cultivating systems that reflect compassion, justice, and human dignity.

Conclusion: Toward a Shared Ethical Future

Responsible innovation in AI demands that we take ethics seriously, not just as a theoretical field, but as a practical guide. As countries like the U.S. strengthen internal governance structures and engage with international partners, there is a growing opportunity to shape a global ethical consensus that is responsive, inclusive, and future-oriented.

By combining thoughtful policy, inclusive design, cultural sensitivity, and ongoing oversight, it is possible to guide the development of AI technologies to uplift rather than undermine human-centered values. The path forward is not easy—but with shared commitment and continuous reflection, it is possible to build a future where AI supports a more equitable, ethical world for all.

(See also Appendix a: “Five Anchors to AI” as a practical approach to implementing ethical standards and guidelines.)

Author (In order of contribution)

Ann M. Marcus, Director, Ethical Tech & Communications, WeAccel
Ann M. Marcus is a Sonoma-raised, Portland-based communications strategist and ethical technology analyst focused on smart cities, community resilience, and public-interest innovation. She leads the Marcus Consulting Group and serves as director of ethical technology and communications at WeAccel.io, a public-good venture advancing mobility, communications, and energy solutions for communities. Ann has advised public and private organizations—including Cisco, the City of San Leandro, Nikon, AT&T, and InfoWorld—on trust-based data exchange, digital public infrastructure, resilience strategy, AI and more. Her current projects include a California senior evacuation program, a Portland robotics hub, and digital energy resource initiatives with utilities in Portland and the Bay Area.

For more information about the Coalition for Innovation, including how you can get involved, please visit coalitionforinnovation.com.

View the Next Chapter