



Responsible AI in Action

Balancing Regulation, Ethics, and the Future

PART 1: Navigating Regulatory Frontiers

Women Defining AI
Community Perspective Paper | January 2024

Responsible AI in Action

Part 1: Navigating Regulatory Frontiers

Responsible AI in Action: Balancing Regulation, Ethics, and the Future is a pivotal paper that serves as a comprehensive guide for the contemporary AI landscape. We are excited to release Part 1 of this insightful series, crafted to empower businesses to make sense of this constantly-changing landscape.

Part 1, ***Navigating Regulatory Frontiers***, provides an in-depth analysis of the current state of AI regulations and their impact on companies, emphasizing significant developments in the European Union, the United States, and the international arena. It sheds light on key regulatory initiatives such as the EU AI Act and the Biden Administration's Executive Order, exploring their implications for various business sectors. This section is crucial for organizations aiming to integrate these new regulatory frameworks into their AI strategies effectively.

While Part 2, ***Ethical AI: Mitigating Risk, Bias, and Harm*** and Part 3, ***Forecast & Takeaways***, will be released subsequently, Part 1 sets the foundation for understanding the critical issues at the intersection of AI innovation and regulation. It provides valuable insights for companies looking to stay ahead in a rapidly changing environment where AI's influence is continuously growing, and the regulatory framework is in constant flux.

Furthermore, because the regulatory landscape is in constant flux, look for periodic updates throughout 2024! Enjoy!

Sincerely,

Women Defining AI



PART 1:

Navigating AI Regulatory Frontiers

AI is expected to be a highly regulated technology but, as it stands, expectations and reality aren't matching up. Few AI-specific regulations actually are in effect. Most reasonably savvy companies understand this, as well as the fact that new laws will start to be enacted across the globe in a fragmented fashion¹. The question they face is how to prepare for this coming unknown without losing competitive edge, particularly at a time when venture capital investment is less certain, the costs of compute and luring engineering talent demand priority, and leadership is juggling countless other issues to keep the company ahead of rapidly accelerating technology.

Our High-Level Recommendations for the Regulatory Landscape

- 1 To access EU markets, prioritize compliance with the EU AI Act
- 2 Less Urgent: The International Community is Still Figuring Things Out
- 3 The US is a Mixed Bag Plus It's an Election Year: Stay Vigilant

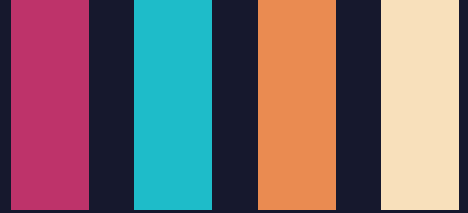
“

If a company moves too fast without regard for consumers or regulations, it can lose its competitiveness and headstart in an instant, and a loss of consumer trust will be difficult to reclaim.

--PART 3 of Responsible AI in Action

THE EU AI ACT

Must Knows



Timing. A draft of the legislative text is expected to be released in January 2024. Assuming the official AI Act is released in April 2024, most of the Act's provisions will be in force by April 2026. Therefore, information in this document may change with ongoing legislative developments.



Application. The AI Act will have broad application. Its regulations will extend to any public or private entity in the world whose AI systems are available on the EU market or affect people located in the EU. However, the AI Act excludes AI systems and models used exclusively for military, national security or defense purposes, or R&D and prototyping activities in pre-release, and also generally excludes free and open source AI systems unless they embody certain risks defined in the AI Act.



Risk-Based Approach. The legal framework of the AI Act reflects a risk-based approach with four categories of AI risk identified: Prohibited AI, High Risk AI, Limited Risk AI, and Minimal Risk AI.



Most Targeted AI: High Risk + Powerful GenAI. The most burdensome obligations under the AI Act apply to High Risk AI and generative AI considered to carry “systemic risks” (generally AI trained with more than 10^{25} FLOPs).



An Enforcement Schedule & Cost for Non-Compliance. The AI Act has a graduated enforcement timeline. While most of the Act's provisions are expected to apply within 24 months of its effective date, EU member states must phase out Prohibited AI within six months, and generative AI regulations will take force within 12 months, of the Act's effective date.

THE EU AI ACT

Must Knows: What are the High Risk & Powerful GenAIs?

“Systemic Risk” GenAI

The European Commission has confirmed that OpenAI’s GPT-4 and “likely” Google DeepMind’s Gemini are the only models to date that qualify under the AI Act as generative AI carrying systemic risks ($>10^{25}$ FLOPS). This will change as more models are trained with greater computing power. Moreover, the European AI Office (to be created by the European Commission as part of the AI Act) is empowered to adjust the FLOP threshold upward or downward in light of ongoing technological developments.

“High Risk” AI

An AI system is deemed High Risk under the AI Act if it fits into one of three buckets:

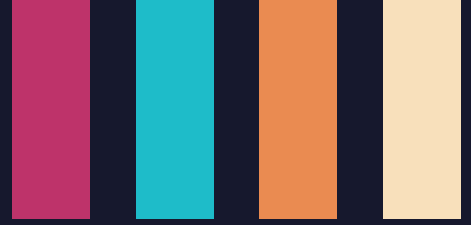
1. The AI is a product already subject to EU product safety legislation (described in the Act’s Annex II) or is intended as a safety component for such products.
2. The AI, or the product incorporating it, already is required to undergo a third-party conformity assessment in accordance with applicable EU legislation, again as set forth in Annex II.
3. The AI is the type explicitly listed in Annex III of the Act. These include AI systems deployed in education or vocational training, for employee recruitment, or in connection with access to essential private and public services. An AI system will always be considered High Risk if it performs profiling of natural persons.



The European Commission has acknowledged that the “vast majority” of AI systems used in the EU now or likely in the future will be considered **Minimum Risk** under the AI Act.

THE EU AI ACT

Nice to Knows



Obligations for High Risk AI Providers

As mentioned, providers of High Risk AI face greater compliance requirements. Specifically, they must subject their technology to a pre-release conformity assessment to show that it complies with the AI Act's detailed requirements for trustworthy AI (e.g. robust criteria around data quality, documentation and traceability, transparency, human oversight, accuracy, log-keeping, and cybersecurity). This assessment has to be repeated any time the system or its purpose are substantially modified.

High Risk AI providers also will have to implement internal quality and risk management systems. Where applicable, and as set forth in the AI Act, providers and deployers of High Risk AI will be subject to specific EU registration requirements - as will providers of AI systems which the provider has deemed not to be High Risk. Moreover, High Risk AI providers located outside the EU will need to appoint an authorized representative in the EU responsible for ensuring the provider's compliance with the AI Act.

Obligations for Developers of GenAI (with and without Systemic Risk)

The scrutiny of generative AI providers is no less rigorous, particularly if they are deemed to carry systemic risks (more than 10^{25} FLOPs). The AI Act requires any generative AI company to respect EU copyright laws when training models, and to make transparent disclosures downstream to their customers and users. Providers of any generative AI also must prepare and keep up-to-date technical documentation of their model, including its training and testing process and the results of its evaluation.

Further, such providers will need to draw up and make publicly available a sufficiently detailed summary about the content used for training of its generative AI model.

Developers of generative AI carrying systemic risk have the added expectation of working with the European AI Office to establish internal codes of conduct, and are under mandate to assess and mitigate risks, report serious incidents, conduct state-of-the-art tests and model evaluations, and ensure adequate cybersecurity.



Providers of High Risk AI are not alone. The AI Act also sets forth obligations for the users, importers, and deployers of High Risk AI.

Nice to Knows

1

UK AISI Announced
At Bletchley, the UK launched the UK AI Safety Institute whose mission is to minimize surprise to rapid and unexpected advances in AI.

2

Let's Keep In Touch!
Additional AI safety summits are planned, with Korea and France to host separate events over the next twelve months.

3

UK Will Not Regulate
After grand-marching this first global AI conference, the UK subsequently made clear that it will refrain from regulating the British AI sector and not introduce any formal AI law in the “short term.”³

The Bletchley Declaration

MUST KNOWS on the International Scene

The UK government convened the first global AI Safety Summit in early November 2023 at Bletchley Park, the famed site where WWII codebreakers like Alan Turing worked. The multilateral talks resulted in the Bletchley Declaration. Endorsed by 28 governments, including those from the US, China, India, Japan and the EU^{1 2}, the agreement acknowledges that “[m]any risks arising from AI are inherently international in nature, and so are best addressed through international cooperation.”

Non-binding. The Bletchley Declaration is non-binding and not legally enforceable.

Emphasis on Frontier AI Companies not SMBs. The Bletchley Declaration largely focuses on “frontier” AI, which is identified as highly capable generative AI models, including foundation models, that can perform a wide variety of tasks. Several developers of such models, including OpenAI, Google DeepMind, Microsoft, and Meta are reported to have attended the summit.

Transparency Encouraged. Companies developing frontier AI models have a heightened responsibility to ensure the safety of their technology. They are “encouraged” to “provide context-appropriate transparency and accountability on their plans to measure, monitor and mitigate potentially harmful capabilities and the associated effects that may emerge, in particular to prevent misuse and issues of control, and the amplification of other risks.”

Pledge to Share Info and Collaborate. Summit attendees resolved to support an internationally inclusive network of scientific research on frontier AI safety, and to collaborate on building risk-based AI policies in their respective countries.



BIDEN'S EXECUTIVE ORDER



Must Knows

In the U.S., President Biden's October 2023 Executive Order (EO) directs more concrete action than the Bletchley Declaration but, like most presidential executive orders, is more guidance than enforceable law. Though a gridlocked Congress and the upcoming presidential election are wild cards in how and when any federal AI regulation may take form, some AI companies and investors are using the EO and its principles as a baseline for shaping their strategies.



Some of the EO is Law; Most is Not. Only one part of the EO has the force of law. The catch is that it doesn't apply to any current AI models . . . yet.



US Agencies are Directed to Act. Various government agencies are directed to propose guidance or regulations addressing a range of AI-related concerns, including consumer protection, antitrust, civil rights, education, financial opportunities, transportation, and healthcare.

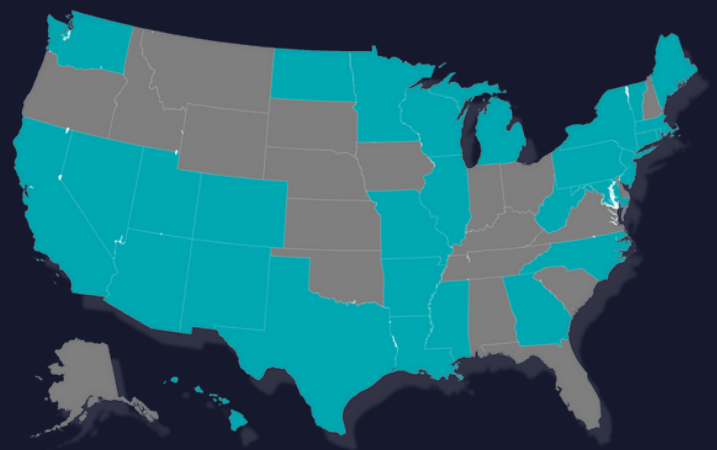


Prioritizing and Growing Technical Talent. The EO wants to boost technical talent in the US. It directs the Department of Homeland Security to streamline the visa and immigration processes for AI experts and AI startup founders, and to use its authority to attract foreign nationals with special AI and engineering skills



Individual States are Starting to Weigh In.

Though a state-by-state analysis of AI-related legislation is beyond the scope of this paper, local legislatures have begun to act. ⁴



No AI Legislation Attempts ⁵

BIDEN'S EXECUTIVE ORDER

Must Knows

The One Part That is Law

Section 4.2 of the EO invokes the Defense Production Act, which has the force of law, to require developers of large scale AI models that could potentially pose a threat to national security, economic security, or health and safety to report on their training (including cybersecurity measures adopted) and red-teaming safety testing results. Set to take effect on January 28, 2024, and subject to final rules from the Commerce Department that could change things, this requirement currently applies to models trained using a quantity of computing power more than 10^{26} FLOPs. This not only is higher than the threshold used in the EU AI Act to help define generative AI carrying systemic risk, but by default exempts existing foundational models (none currently are that large).



Agency Direction Too Ambitious?

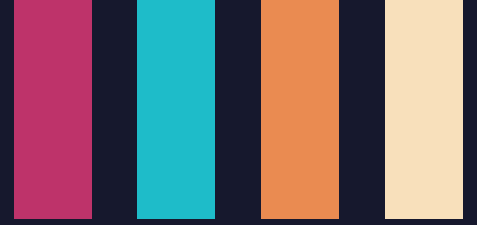
It remains to be seen whether the U.S. agencies, which include the Commerce Department, FTC, and the National Institute of Standards and Technology (NIST), will be able to meet the aggressive deadlines set forth in the EO and whether they have the necessary budget and resources available.

Thus far, the FTC has been very proactive - issuing a series of AI-related guidance beginning in 2020 and, in 2023, launching a broadside investigation into OpenAI's practices and banning the pharmacy chain RiteAid from using AI-powered facial recognition technology for five years, making clear that their existing legal enforcement powers extend to AI.



BIDEN'S EXECUTIVE ORDER

Nice to Knows



What's Expected of the Copyright Office + Homeland Security?

The EO directs the Copyright Office to publish a report and issue recommendations on potential executive actions relating to copyright and AI. Homeland Security is ordered to develop a program to mitigate AI-related intellectual property risks by identifying AI-related intellectual property theft and assisting the private sector with mitigating intellectual property theft violations.

What's Expected of the NIST?

The EO directs the NIST to develop guidelines and best practices for the secure development, evaluation and auditing, red teaming, and ensuring that AI models are safe, secure, and trustworthy. NIST also must create a companion resource to its AI Risk Management framework for generative AI. Shortly after the EO issued, the administration announced the creation of the United States AI Safety Institute (US AISI), which will sit in NIST and is tasked with operationalizing the framework. NIST also is tapped to develop standards on privacy and on authenticating when content is AI-generated. The US AISI is expected to develop technical guidance for regulators directed to propose rules pursuant to the EO, and to collaborate with the UK AISI announced at Bletchley.

What's Expected of the Bureau of Industry and Security? BIS will propose rules regarding the reporting requirements for frontier AI models discussed above that are based on the EO's invocation of the Defense Production Act.

Cybersecurity Mentions

Per the EO, the Treasury Secretary must issue a public report on best practices for financial institutions to manage AI-specific cybersecurity risks. The Secretary of Homeland Security will evaluate how deploying AI may make critical infrastructure systems more vulnerable to critical failures, physical attacks, and cyber-attacks, and create an advisory committee to advise on improving security.




Absent any overarching federal law, AI regulation in the US is poised to go the way of data privacy law and become a patchwork of different rules and enforcement frameworks for companies to figure out.

NEXT STEPS

Buckle up - lots will happen this year!

THE EU AI ACT	<p>AI developers already accessing the EU market or whose technology affects people living in the EU, or that are planning to launch in any EU countries, will need to determine their risk profile under the AI Act and relevant compliance obligations.</p> <p>Companies that use AI systems in the EU as a customer should review their vendor agreements to ensure they contain appropriate representations and warranties in compliance with the AI Act, as well as reasonable indemnification for the vendor's noncompliance. Deployers and importers of High Risk AI also will need to develop an EU compliance program.</p> <p>Don't forget about the GDPR (General Data Protection Regulation)! Preparedness around the AI Act is only one piece of the puzzle for accessing EU markets. GDPR compliance is the other, and data protection authorities are becoming active in the AI space.</p>
THE BLETCHLEY DECLARATION	<p>The power of this declaration is that it happened at all. Companies should be aware of this symbolic first step towards international cooperation on AI policy, and the unified areas of concern coming into focus.</p> <p>Watch to see how the Bletchley Declaration's words may eventually translate into action. Already we've seen the US follow the UK and create its own AI Safety Institute. But we've also seen the UK hedge on promulgating actual legislation.</p>
BIDEN'S EXECUTIVE ORDER	<p>Be on the lookout for new guidance and proposed agency regulations stemming from the EO. Expect the Commerce Department (including NIST, BIS, and the newly created US AISI) and the FTC to be especially active on issues relevant to business.</p> <p>Partnerships between civil society groups, venture capitalists, and government are starting to emerge. An example is <u>Responsible Innovation Labs</u> which, in collaboration with the Commerce Department, is securing <u>voluntary commitments from venture funds</u> around how the startups they back should develop AI responsibly. This is in step with the White House's campaign to obtain <u>voluntary commitments from companies</u>.</p>



Closing Thoughts

As we conclude our exploration of *Part 1, Navigating Regulatory Frontiers*, it's evident that the journey of integrating AI into our businesses and society is both exhilarating and complex. The landscape we've traversed in this paper underscores the critical importance of staying agile and informed in a rapidly evolving domain where innovation, ethical considerations, and regulatory frameworks intersect.

We encourage leaders and practitioners in the AI field to absorb the insights from this paper and to prepare for the dynamic future of AI. By proactively adapting to the evolving landscape and integrating ethical and regulatory considerations into AI strategies, businesses can lead the way in creating a future where AI is not only powerful and innovative but also responsible and beneficial for all.

Be on the look out for **Part 2, Ethical AI: Mitigating Risk, Bias, and Harm** and **Part 3, Forecast and Takeaways** coming in February 2024.

Responsible AI in Action

*Balancing Regulation,
Ethics, and the Future*



REFERENCES

¹ For an interactive world map that tracks AI regulatory and policy updates, check out this [Global AI Regulation Tracker](#) by attorney Raymond Sun.

² For a comprehensive analysis of China's AI regulations and their relevance to global public policy, see [China's AI Regulations and How They Get Made](#), Carnegie Endowment for International Peace, Sheehan, M. (2023, July 10)

³ Source: Article in Financial Times: [UK Will Refrain From Regulating AI 'in the short term'](#) (paywall)

⁴ See NYC [Local Law 144](#) and its [Final Rule](#), effective July 5, 2023, which makes it an unlawful employment practice for employers to use automated employment decision tools (AEDTs) to screen candidates and employees within New York City unless certain bias audit and notice requirements are met.

⁵ US Map built with information from [National Conference of State Legislatures](#). Updated January 12, 2024.



About Us

Women Defining AI (WDAI) is a trailblazing organization focused on empowering women and non-binary individuals in artificial intelligence. We offer a unique blend of hands-on learning and community support to engage mid-career individuals with non-technical backgrounds to demystify and ultimately define AI. In our mission to democratize AI knowledge, making it accessible, relatable, and engaging, we aim to be a vital force in shaping the future of women in technology.



[Follow our page](#)



[Join Us!](#)

www.womendefiningai.com



[Email the team:](#)

info@womendefiningai.com

Contributing Authors:



Irene Liu

Founder & Advisor
Hypergrowth GC

Executive in Residence
UC Berkeley School of
Law



Shella Neba

Chief Legal Officer &
Strategist



Teresa Burlison

General Counsel & Advisor

Editor:



Nichole Sterling

Co-founder
Women Defining AI

Disclaimer. The opinions expressed in this Community Perspective reflect solely upon the contributors and not the organizations or companies they are associated with.