

nasscom

Responsible AI

Guidelines for Generative AI



nasscom

This document contains a description of the normative obligations for researchers, developers and users of generative AI models and applications (hereinafter “guidelines”), with references to corresponding socio-technical recommendations and aids, where possible, to help ensure responsible adoption of generative AI.

These guidelines are the result of consultations with a multi-disciplinary group of AI experts, researchers and practitioners, with representations from the industry, academia, and civil society.

Recognising the fast-paced nature of the technological and regulatory developments impacting generative AI practice, we plan on releasing future editions of this document with revisions and updates that could serve stakeholders better over time.



Preamble

The recent powerful developments in generative artificial intelligence (hereinafter "GenAI") and the projected pace of adoption of this revolutionary technology across industries have got people both excited and worried about the future. The excitement comes from the promise of GenAI to massively improve industrial productivity, boost economic growth¹ and scale humanitarian efforts². The worry comes from our sheer unpreparedness³ to effectively tackle the fast-evolving universe of societal harms that are anticipated with the adoption of GenAI⁴.

Stakeholders, largely, do not yet align on a set of robust, common standards and protocols for researching, developing, and using GenAI responsibly. This heightens the risk of

GenAI being deployed without appropriate safeguards in place, impacting public safety and potentially pausing the pace of innovation in GenAI⁵.

At nasscom, we believe this calls for urgent action from all concerned stakeholders to cooperate and co-develop responsible approaches for researching, developing and using generative AI models, applications, and tools in alignment with positive human values, for the benefit of humanity as a whole.

In line with that spirit, we propose this body of guidelines for those engaged in researching, developing, and using generative AI technologies, with the following provisos:

- This document is not an operational manual or guidebook. It is rather intended to build stakeholder consensus on the core normative obligations of those engaged in researching, developing, or using GenAI technologies.
- We expect the guidelines to become instrumental in defining frameworks for the development of standards, protocols, audit checklists, certifications, monitoring and evaluation mechanisms, and other forms of actionable guidance, tools, and best practices to effectively mitigate potential harms from the adoption of GenAI.

Section 1. Object and Scope

The guidelines focus on research, development, and use in relation to GenAI. The guidelines define GenAI as a type of artificial intelligence technology that can create artefacts such as image, text, audio, video, and various forms of multi-modal content.

The object of these guidelines is to promote and facilitate responsible development and use of GenAI solutions by different stakeholders. The guidelines also intend to achieve a robust, common understanding of normative obligations amongst stakeholders to help them improve their net social impact with GenAI and to foster trust in the adoption of GenAI technologies across industries.⁶



Research on GenAI solutions may include both fundamental and applied research on GenAI models, applications, tools and techniques.



Development of GenAI solutions may include building and deploying GenAI models and sourcing training data for GenAI model design and development to design or power various commercial, non-commercial or personal applications.



Use of GenAI solutions may include employing or modifying GenAI models for any product or service development, enhancement or delivery, conducting scientific or academic research, writing computer code, generating creative outputs (e.g., artworks, musical compositions, product designs) for any purpose - commercial, non-commercial, or personal.

Note that the categories of “research”, “development”, and “use” are not mutually exclusive; a given stakeholder could fit into all three categories. Stakeholders may include, but not be limited to, technology companies, startups, open source developers and researchers.

Section 2.

Index of Potential Harms

The guidelines seek to aid mitigation of the following harms associated with the research, development, and use of GenAI technologies:



Proliferation of misinformation, disinformation, hateful (seditious, defamatory, socially disharmonising) content⁷



Infringement of intellectual property⁸ (e.g., copyrighted works of art, patented designs and inventions, trademarks, trade secrets) and academic malpractice⁹ (e.g., plagiarism in research)



Privacy harms through violations of data protection norms and standards¹⁰



Propagation of harmful social, economic, and political biases¹¹



Large-scale job displacements, and loss of livelihood and economic strain for a considerable portion of the existing industrial workforce¹²



Huge carbon and water footprints and associated environmental degradation¹³



Surge in malicious cyber attacks¹⁴

Section 3.

Obligations of those conducting fundamental and applied research on GenAI models, applications, tools, and techniques (hereinafter “researchers”)¹⁵

**A**

Demonstrate reasonable caution and foresight by systematically and rigorously anticipating and evaluating both positive and negative contingencies that might arise from the conduct of research using techniques like horizon scanning, scenario planning, etc.

**B**

Demonstrate transparency and accountability by releasing public disclosures about the values, goals, and motivations for driving or funding a research project and by describing the methodologies, model training datasets, and tools adopted for the conduct of research in all such disclosures.

**C**

Demonstrate reliability and safety by adhering to established privacy-preserving norms and standards in research data collection, processing, and usage, and conducting safety testing of GenAI models in regulated environments.

**D**

Demonstrate inclusion by accounting for the risk of harmful bias in research and deploying protocols and measures to mitigate it, and by publishing research findings in open-source formats, wherever possible, to democratise framing of new problem statements to advance the state-of-the-art in GenAI, foster collective inquiry into the potential risks and benefits from the adoption of GenAI technologies, and engender prevalent societal values in GenAI.

**E**

Support progress of humanity as a whole by prioritising research on GenAI applications, tools, and techniques that hold the maximum potential to enhance human agency and improve the human condition, and by advancing research in technical AI safety to aid the development of best practices for developing, deploying, and using GenAI in a safe, trustworthy, and environment-friendly manner.

Section 4.

**Obligations of those engaged in the development of
GenAI solutions (hereinafter "developers")¹⁶**

A

Demonstrate reasonable caution and foresight by evaluating potential harms from the development, deployment, and use of a GenAI solution through comprehensive risk assessments and internal oversight throughout the lifecycle of the solution, with the aid of critical guiding frameworks (e.g., [nasscom's Responsible AI Governance Framework](#), [NITI Aayog's Approach for Operationalising Principles for Responsible AI](#), [UNESCO's Recommendation on the Ethics of AI](#), [OECD AI Principles](#)), by retaining qualified personnel (e.g., risk and compliance officers, ethics councils), and by prescribing terms of service and guidance for safe and responsible use by private individuals and downstream model and application developers¹⁷, while disclosing questionable uses of the solution.

B

Demonstrate transparency by publicly disclosing the data and algorithm sources used for modelling and all other technical, non-proprietary information about the solution's development process, capabilities and limitations¹⁸. The disclosure of any or all such information may be withheld only if there is reasonable apprehension amongst the developers that such disclosure would risk malicious use of the solution and imminent harm to public safety. Developers shall prove such reasonable apprehension to the satisfaction of the regulator whose jurisdiction applies.

C

Demonstrate reliability and safety by practising due diligence in the adoption of means and methods for solution development and deployment (e.g., [nasscom's Responsible AI Architect's Guide](#)) through strict adherence to applicable data protection and intellectual property rules and norms in the collection, processing and usage of model training data¹⁹, industry best practices for model design, development, deployment, and monitoring (e.g., feeding contextual awareness in model design, designing and development of models by diverse and multidisciplinary teams, phased deployment of large-scale consumer facing applications, human in the loop design, fair and lawful interpretation of results generated by the solution), and deployment of best-in-class cybersecurity infrastructure.

 D

Demonstrate accountability by devising technical means for furnishing explanations for outputs generated by GenAI solutions in high-stake contexts (e.g., use of GenAI in consumer credit lending) in accessible, intelligible forms, and by establishing and adhering to fair and equitable schemes for liability apportionment and mechanisms for grievance redressal to deal with mishaps caused by the development, deployment, and use of GenAI solutions.

 E

Support progress of humanity as a whole by developing, deploying, and using GenAI solutions in alignment with the positive goals of human progress and societal well-being, in environment-friendly ways that prioritise energy efficiency in model training and development, and by disclosing the solution source code, wherever possible, to reduce duplication of efforts and make computing overall more efficient and sustainable for all.



Section 5.

**Obligations of those using GenAI solutions for commercial,
non-commercial, or personal purposes**

(hereinafter "users")²⁰

A

Demonstrate reasonable caution and foresight and support progress of humanity as a whole through internal oversight and comprehensive risk assessments covering the entire lifecycle of the GenAI solution, with the aid of critical guiding frameworks (e.g., [nasscom's Responsible AI Governance Framework](#), [NITI Aayog's Approach for Operationalising Principles for Responsible AI](#), [UNESCO's Recommendation on the Ethics of AI](#), [OECD AI Principles](#)), by retaining qualified personnel (e.g., risk and compliance officers, ethics councils), by prescribing terms of service and guidance for safe and responsible use by private individuals and downstream model and application developers²¹, while disclosing questionable uses of the solution, and by balancing any substitution of the current workforce that is induced by the solution's adoption with proportionate investments in worker upskilling and reskilling programs across critical business roles and functions²².

B

Demonstrate transparency by publicly disclosing all technical, non-proprietary information about the development process, capabilities and limitations of the downstream models and applications (including *inter alia* details of any modifications performed on the models and corresponding data and algorithm sources), the use of the GenAI models or features for product or service delivery or enhancement, in deliverables generated within an academic or commercial setting, in production of research outputs²³, and in claims for intellectual property²⁴.

C

Demonstrate reliability and safety by using the GenAI solution in compliance with the terms of service, prescribed guidance and applicable public regulations for safe and responsible use, by developing and deploying downstream models and applications in compliance with industry best practices (e.g., feeding contextual awareness in model design, development of applications by diverse and multidisciplinary teams, phased deployment of large-scale consumer facing applications, human in the loop design, deployment of applications with guidance for responsible

use, fair and lawful interpretation of results generated by the solution), by not using the solution to infringe or misappropriate the rights of others or to propagate disinformation or harmful social, economic, or political biases, by exercising caution or due diligence in consuming or using content or media generated by the solution, by refraining from sharing any sensitive, confidential, or personally identifiable information in the course of interaction with the solution, and by instating corporate and institutional safeguards and codes of conduct to counter misuse or unauthorised use of the solution²⁵.



Demonstrate accountability by devising technical means for furnishing explanations for outputs generated by GenAI solutions in high-stake contexts (e.g., use of GenAI in consumer credit lending) in accessible, intelligible forms, and by establishing and adhering to fair and equitable schemes for liability apportionment and mechanisms for grievance redressal to deal with mishaps caused by the use of GenAI solutions.



Support progress of humanity as a whole by using GenAI solutions in alignment with the positive goals of human progress and societal well-being, in environment-friendly ways that prioritise energy efficiency in model redesigning and use, and by disclosing the solution source code, wherever possible, to reduce duplication of efforts and make computing overall more efficient and sustainable for all.

Section 6.

Joint Obligation of researchers, developers, and users



Support progress of humanity as a whole

by committing to steer and support universal AI literacy and awareness programs, technical AI safety research, and regulatory reform projects focused on designing guardrails for safe and trustworthy adoption of GenAI technologies for the benefit of all stakeholders, in collaboration with governmental and non-governmental organisations.

Section 7.

Compliance with Guidelines

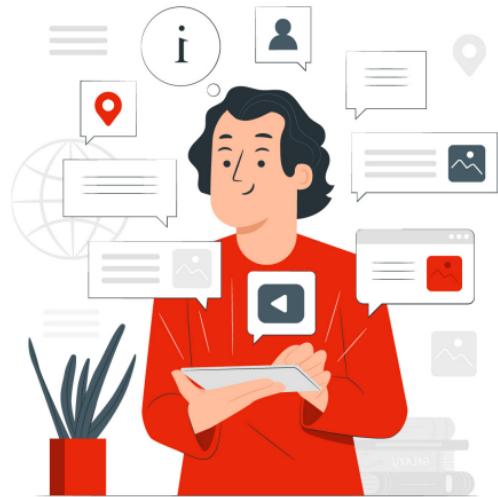


As researchers, developers, and users of GenAI technologies, we resolve to contribute to the co-development and adoption of actionable guidance, tools, and best practices to enable all stakeholders to successfully comply with the guidelines and help humanity advance with the safe and responsible use of GenAI.

References

#GenerativeAI

1. Generative AI could raise global GDP by 7% (Goldman Sachs Research), [accessible here](#).
2. Mahanand, Jay, Generative AI: A game-changer for humanitarian assistance (United Nations World Food Programme), [accessible here](#).
3. Hacker, Philipp et al. Regulating ChatGPT and Other Large Generative AI Models (Faculty of Law Blogs, University of Oxford) 1 March 2023, [accessible here](#).
4. Lee, Michelle, and Kruger, Lucas, Risks and ethical considerations of generative AI (Deloitte) 13 March 2023, [accessible here](#).
5. Pause Giant AI Experiments: An Open Letter (Future of Life Institute) 22 March 2023, [accessible here](#). On 30 March 2023 US Federal Trade Commission (FTC) received a submission from the Center for Artificial Intelligence and Digital Policy (CAIDP) requesting it to investigate OpenAI's practices and bar the company from issuing new commercial releases of GPT-4, see CAIDP's submission, [accessible here](#).
6. Longoni et al. News from Generative Artificial Intelligence is Believed Less (ACM Conference on Fairness, Accountability, and Transparency (FAccT '22), June 21–24, 2022, [accessible here](#). Trust in Artificial Intelligence: A Global Study 2023 (KPMG and The University of Queensland, Australia) 22 February 2023, [accessible here](#).
7. Goldstein, Josh, Generative AI Is Enabling Fraud and Misinformation: Here Is What You Should Know (Center for Security and Emerging Technology, Georgetown University), 17 January 2023, [accessible here](#). Giansiracusa, Noah and Marcus, Gary, Big Tech Hasn't Fixed AI's Misinformation Problem Yet (Time Magazine) 13 February 2023, [accessible here](#).
8. Appel et al. Generative AI Has an Intellectual Property Problem (Harvard Business Review) 7 April 2023, [accessible here](#).
9. Westfall, Chris, Educators Battle Plagiarism As 89% Of Students Admit To Using ChatGPT For Homework (Forbes) 28 January 2023, [accessible here](#).
10. Burgess, Matt, ChatGPT Has a Big Privacy Problem (WIRED) 4 April 2023, [accessible here](#). Koerner, Katharina, Generative AI: Privacy and tech perspectives (International Association of Privacy Professionals), [accessible here](#).
11. Gordon, Rachel, Large language models are biased: Can logic help save them? (MIT News) 3 March 2023, [accessible here](#).
12. Briggisi et al. The Potentially Large Effects of Artificial Intelligence on Economic Growth (Economics Research, Goldman Sachs) 26 March 2023, [accessible here](#). Mok, Aaron and Zinkula, Jacob, ChatGPT may be coming for our jobs. 2 February 2023, [accessible here](#).
13. Stokel-Walker, Chris, The Generative AI Race Has a Dirty Secret (WIRED UK) [accessible here](#).
14. Dor, Dorit, It's up to us to determine if generative AI helps or harms our world (World Economic Forum), [accessible here](#). OpenAI: Cybercriminals starting to use ChatGPT (Check Point Research) 6 January 2023, [accessible here](#).
15. Stilgoe, Jack, Developing a Framework for Responsible Innovation (Elsevier) November 2013, [accessible here](#).
16. Early Thoughts on Generative AI: Prepared Remarks of Commissioner Alvaro M. Bedoya, Federal Trade Commission Before the International Association of Privacy Professionals (Federal Trade Commission, Washington D.C. United States of America) 5 April 2023, [accessible here](#).
17. Engler, Alex, Early Thoughts on Regulating Generative AI like ChatGPT (Brookings Institution) 21 February 2023, [accessible here](#).
18. However, note that OpenAI recently claimed an exception to this obligation citing competition and public safety considerations, [accessible here](#).
19. Note that OpenAI in their recent response to a Data Subject Access Request defended collecting large amounts of publicly available personal data on grounds of "legitimate interests", [accessible here](#).
20. Bouzid, Ahmed, Break the AI Blame Cycle: User Responsibility in the Age of Generative AI (CMSSWire) 10 April 2023, [accessible here](#).
21. Engler, Alex, Early Thoughts on Regulating Generative AI like ChatGPT (Brookings Institution) 21 February 2023, [accessible here](#).
22. Wright, Scott A. and Schultz, Ainslie E. The rising tide of artificial intelligence and business automation: Developing an ethical framework (Elsevier) November-December 2018, [accessible here](#).
23. Brainard Jeffrey, As scientists explore AI-written text, journals hammer out policies (Science) 22 February 2023, [accessible here](#).
24. Belanger, Ashley, Authors risk losing copyright if AI content is not disclosed, US guidance says (ArsTechnica) 17 March 2023, [accessible here](#).
25. Sloan, Karen, University of California Berkeley Law School Rolls out AI policy ahead of final exams (Reuters) 21 April 2023, [accessible here](#).



Acknowledgements

#GenerativeAI

We would like to thank all our reviewers for sharing their expert feedback and inputs for overall guiding the preparatory phase of this document to ensure that it met the highest ethical standards.

With their generous and timely support, we were able to demonstrate industry's much-needed steadfastness in responding to what remains one of the toughest regulatory issues of our times.



nasscom

Write to us at **responsibleai@nasscom.in**

© nasscom