International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 19-26 | www.ijaiet.com |

The Social Impact of Generative AI: A Framework for Assessing Misinformation, Labor, and Equity Risks

Dr. Sandeep Kumar

1Department Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India City- Nagpur, Country-India, Pincode-440008

*Corresponding author
Dr. Sandeep Kumar
Symbiosis Institute of Technology, Nagpur Campus, Symbiosis International (Deemed University), Pune, India City-Nagpur, Country-India, Pincode-440008
E-Mail: er.sandeepsahratia@gmail.com Orcid : https://orcid.org/0000-0002-4752-7884

Received : 13-06-2026 | Revised : 20-08-2026 | Accepted : 27-08-2026 | Published : 31-08-2026

Abstract

The development and proliferation of generative artificial intelligence systems that can generate text, images, audio, and code on scale has spread into society in an unprecedented way, and has given rise to a heavily fragmented debate on the social implications of this technology. That speech is usually either a list of harms or a list of benefits; it doesn't often present a systematic method to evaluate particular risks in particular contexts. While this paper introduces a conceptual framework for social impact assessment of generative AI, it focuses on three areas where impacts are most significant and where they are most discussed: misinformation, labor, and equity. The framework defines the principal risks and the mechanisms that create them for each domain and a common set of cross-cutting dimensions (likelihood, severity, reversibility, distribution, timescale, and tractability) to which any candidate risk can be brought to evaluate it. It combines these with domain-specific indicators to anchor assessment in observable evidence, a matrix between the level of severity and likelihood to establish mitigation priority and a mapping of governance and mitigation levers to actors placed to act on them. The framework deliberately aims to be evaluative as opposed to alarmist or dismissive; it is meant to balance potential risks and benefits and to differentiate between potential and actual harms in the three domains, acknowledging vastly different empirical evidence from one to another. It brings a common language and a replicable process for researchers, developers and policy makers to get beyond the general worry or excitement about generative AI to specific evidence-based and comparable evaluations of the social risks of these tools.

Keywords: generative AI; social impact; misinformation; labor displacement; algorithmic equity; AI governance; risk assessment framework

1. Introduction

In just a few years, the ability of generative artificial intelligence (AI) models to generate fluent text, realistic images, audio, and functional code on command has progressed from the realm of research novelty to becoming a daily tool. As it diffused, so too have the claims about its social implications followed, from warnings of an epistemic crisis and mass unemployment to promises of widespread productivity gains and democratised access to expertise. The

discourse is, however, fractured, and often polarized – it can move from listing potential harms without estimating the probability and/or severity of these consequences, or from focusing on benefits to considering costs as secondary [1], [2].

What's missing is a systematic method for determining the likelihood of a specific risk in a specific setting, and what the severity of the risk is, who would be affected, and whether it can be reduced or addressed. There is a need for an applied assessment procedure that will link the identification of a risk to its evaluation, and to the governance response it requires, as existing risk taxonomies have mapped the landscape of possible harms with excellent value and ethics frameworks have articulated high-level principles [2], [3], [4].

This paper provides such a framework, and has been shaped by three areas in which the social stakes of generative AI are seen as highest, and debate is most vibrant: misinformation – use of generative systems to create and disseminate misinformation, false, or misleading content; labour – impact of automation and augmentation on work, skills and livelihoods; and equity – inequitable distribution of harms, benefits, and representation across groups. The domains selected are not intended to be exhaustive, as they represent the social impact of generative AI, but rather serve as illustrative examples of social impacts that are consequential, contested, and where the framework is designed to provide assessment.

The framework is designed to be both evaluative and not alarmist or dismissive. It is designed to balance risk and benefit and to differentiate between proven and potential harm, where the evidence in the 3 domains is very unevenly distributed. It provides a structured account of the three domains and their mechanisms (Figure 1 and Table 1); a common set of cross-cutting assessment dimensions and a matrix for the prioritisation of risks (Table 2 and Figure 2); a set of domain specific indicators to ground assessment in evidence (Table 3); a mapping of governance levers to responsible actors (Table 4). The paper is conceptual, it does not report new empirical measurements of harm, but rather establishes a procedure for assessing harm.

2. Background

2.1 Generative AI and its social salience. Whereas previous AI generated work was classification or prediction, generative models produce open-ended artifacts, such as prose, images, speech, and code, they're scalable, they're cheap to produce, and they're fluent and can be hard to tell apart from human output. It is these qualities that make their social impact so unique—those characteristics that can help a student or a clinician can also produce a plausible news report, a suite of repetitive tasks, or replicate societal biases on a massive scale [1]. Thus, social impact cannot be deduced from the technology itself, but rather from the context, people and deployment of a capability.

The sociotechnical nature implies a methodological implication. General evaluations of the technology often lead to statements that are both true and unhelpful, because there's no way to tell if any particular system will create the harm, and because they can't help guide decisions about any system. Assessment, on the other hand, should be linked to a context of use: a specific model, application, population and setting. The framework here devised is based on this idea that social impact is not inherent in the model itself and that it is a property of a deployment.

2.2 The case for a structured framework. There are two features of the current discourse that motivate a structured approach.There are two features of the current discourse that call for a structured approach. Firstly, risks are often defined without any reference to their level of calibration, which in some cases brings them to the level of certainty when they are not, or are cast as science fiction when they are actually happening or are likely to happen soon, so there is a need for a uniform way to describe the likelihood and severity of risks. Second, the three domains are very unevenly substantiated: some equity harms are already empirically substantiated, some labor effects are now being measured, while some misinformation harms are largely prospective. The assessment framework should allow for this irregularity to not be characterized as a proven or hypothesized risk.The assessment framework should be flexible enough to not only include a demonstrated risk but also a speculative risk [2].

2.3 Weighing risks against benefits. A credible assessment framework must also acknowledge the real benefits that generative AI brings with respect to supporting routine work, broadening access to expertise, and aiding creation and analysis, and move beyond simply counting risks to measuring and valuing the net social impact. The framework below is intended to be part of that "benefit-aware" assessment that precedes the identification and prioritisation of

International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 19-26 | www.ijaiet.com |

risks, and that must be addressed before reducing them without sacrificing the benefits that drive deployment to begin with.

3. The Assessment Framework

3.1 Structure and overview. Figure 1 illustrates the framework, which starts with capability, moves to consequence, and ends with response. Each of the three domains poses risks that can be generated by Generative-AI, with a common set of dimensions across domains for assessment; the assessment delivers a priority rating; and the priority rating drives a governance response. The results of deployment and mitigation are fed back into the assessment as they will shape the social impact of generative AI, as will its capabilities, adoption, and countermeasures. The three domains and their main risks are outlined in Table 1.

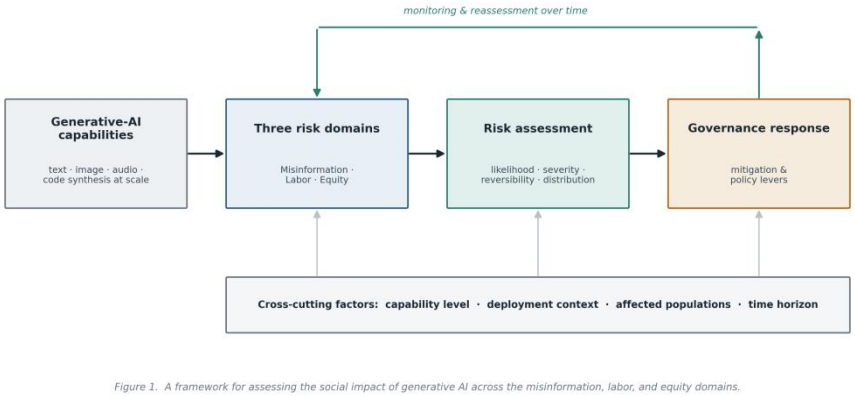

Figure 1. *A framework for assessing the social impact of generative AI across the misinformation, labor, and equity domains.*

Table 1. *The three social-impact domains, their principal risks, and illustrative generating mechanisms.*

Domain	Principal risks	Illustrative mechanisms
Misinformation	Synthetic disinformation and deepfakes; scaled influence operations; erosion of epistemic trust	Cheap high-quality generation; micro-targeting; sheer volume overwhelming verification
Labor	Task automation and displacement; task reshaping and skill devaluation; effects on wages and job quality	Substitution of routine cognitive tasks; shifts in the demand for skills; augmentation of some workers
Equity	Representational and allocational bias; unequal access and benefit; linguistic and cultural exclusion; concentration of power	Bias in training data; uneven performance across groups; asymmetric access to models and compute

3.2 The three risk domains. For misinformation, generative models reduce the cost and increase the quality of content they can produce, allowing for widely accessible, easy-to-produce, believable synthetic text, images and video content, and enabling influence operations at scale [5], [6], [7]. The capacity is not challenged, it is just a question of impact. While experimental evidence shows that AI-generated text can be perceived as credible [8], and that synthetic political video can cause uncertainty [9], there is limited and contested evidence of persuasion and real-world behavioural impact at large scale, and a careful assessment treats large-scale harm as plausible but not yet demonstrated. In the labor domain, the direction of the overall impact of the LLM can even be ambiguous, though analysis of task exposure indicates a significant portion of work tasks could be impacted [10], [11]. A displacement view focuses on cognitive task substitution, a complementarity view on cognitive task augmentation and the creation of new tasks [12]; field experiments show substantial productivity gains, with some reporting disproportionate gains

International Journal of Artificial Intelligence and EvolveTech
ISSN: 0000-0000
Volume 1, Issue 1 | PP 19-26 | www.ijaiet.com |

for less experienced workers, which may contribute to both reducing and expanding inequality [13], [14]. However, some harms in the equity domain are relatively well documented: harms to society can be reproduced in models, accuracy differences have been demonstrated across groups in commercial systems, and benefits and governance of the technology is unevenly distributed across language, region and community [15], [16], [17], [18]. When it comes to assessing risk, it's not just about what might happen, but how solid each concern is.

These are not separate domains and a full assessment takes into account the combination of risks across domains. Labor effects have a non-uniform impact on the harm of misinformation, partly because of equity concerns, as the distribution of the underlying capability is unevenly distributed across communities, and because of the way the labour effects relate to existing advantage, as well as the fact that the distribution of the underlying capability is unevenly distributed across communities, which affects all three. The cross-cutting distribution dimension added below is in part intended to bring these interrelationships to the fore, making this moderate risk not be overlooked when the risk effects are focused in the same population.

3.3 Cross-cutting assessment dimensions. The framework compares these disparate risks in six ways, identified in Table 2, which are the probability that the harm will be realised, the magnitude of the harm if it does happen, whether those affected can recover or not, who bears the risk relative to who benefits, whether the harm is immediate or long run, and whether there are mitigation options available. The second and third dimensions can be combined to form the severity-by-likelihood matrix of Figure 2, a typical risk-assessment tool that assigns each identified risk to a level of concern and assigns a mitigation priority based on the likelihood and severity; the last dimension is the additional qualification that irreversible or inequitably distributed harm has greater concern than the likelihood and severity alone would suggest. The matrix is not a tool for judgment, but an alternative to it; the assignment of any particular risk should be based on evidence from the domain and not be asserted in the matrix.

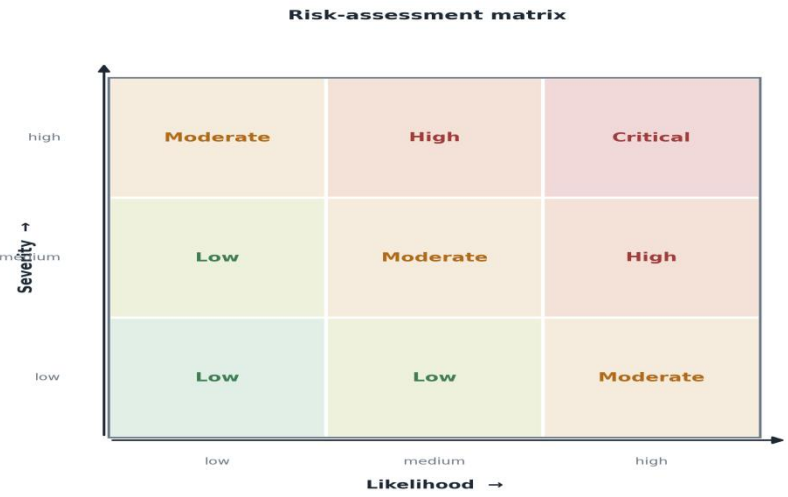

Figure 2. *A severity-by-likelihood matrix onto which each assessed risk is placed to set mitigation priority.*

Table 2. *Cross-cutting dimensions for assessing any candidate social-impact risk.*

Dimension	What it captures	Guiding question
Likelihood	Probability the harm materialises	How plausible is it, given the capability and how it is deployed?
Severity	Magnitude of harm if it occurs	How many are affected, how badly, and at what scale?
Reversibility	Whether the harm can be undone	Can affected parties recover, or is the damage lasting?
Distribution	Who bears risk versus benefit	Are burdens and benefits shared fairly across groups?

Dimension	What it captures	Guiding question
Timescale	When the harm manifests	Is it immediate, near-term, or long-run?
Tractability	How amenable to mitigation	Can available levers meaningfully reduce it?

3.4 Indicators and governance responses. Assessment is only as useful as the evidence that supports it, so the framework correlates each domain with the indicators listed in Table 3 that shift judgements from speculation to measurement in the prevalence and detectability of synthetic content, and in measured prevalence and effect on beliefs in the case of misinformation, task-exposure, productivity, and employment measures in the case of labour, and disaggregated performance audits and access analyses in the case of equity. Assessment then leads to a governance response once completed. Table 4 assigns mitigation levers to the actors that can apply them, in recognition that there is a distributed responsibility, from developers to build-in provenance and auditing, and data practices be made more inclusive; from platforms and firms, to adopt policies and redesign tasks; from governments, to support the transition, standards, and oversight; from affected communities, to participate in design and evaluation. There is no single player who holds all the levers; it usually takes several players to be effective in mitigation.

Table 3. *Domain-specific indicators that ground the assessment in observable evidence.*

Domain	Assessment indicator	Example evidence source
Misinformation	Prevalence and detectability of synthetic content; measured effect on beliefs and trust	Platform data; controlled experiments; fact-checking corpora
Labor	Task exposure; productivity and employment effects; wage and job-quality changes	Task-level analyses; field experiments; labor-market statistics
Equity	Performance disparities across groups; distribution of access and benefit	Disaggregated model audits; access surveys; representation analyses

Table 4. *Governance and mitigation levers mapped to domains and the actors positioned to apply them.*

Domain	Response and mitigation levers	Primary actors
Misinformation	Content provenance and watermarking; detection; platform policy; media literacy	Developers, platforms, regulators, educators
Labor	Reskilling and transition support; task redesign; labor and education policy	Firms, governments, educators, workers
Equity	Bias auditing; inclusive data and design; access programs; participatory evaluation	Developers, auditors, affected communities, regulators
Cross-cutting	Transparency, impact assessment, standards, and ongoing monitoring	All actors, coordinated

4. Discussion

4.1 Using the framework. The framework is designed to be used in a specific context and with a specific system and not generalized to generative AI in general. The candidate risks for each of the domains are identified, and the indicators that are relevant to those risks are collected, evaluated to the six dimensions, then placed on the severity-by-likelihood matrix, and a prioritised set of governance responses are read off. It's less important what they think of it than it is that they impose a consistent and transparent framework for their discussion for example, if two analysts disagree on a risk, they can pinpoint it more specifically in an estimate of likelihood, a weighting of distribution or a reading of the evidence, rather than talking past each other.

4.2 The uneven evidence base. One of the core principles of using the framework in all three domains is that they are very unevenly developed in terms of the evidence that they contain; and that the way you assess them responsibly needs to reflect that. Several equity harms are illustrated and can be measured based on evidence; the effects on labor are actively measured, and we would expect net effects to depend on distributional choices, and the greatest misinformation harms are still possible but not yet established at scale. Combining them all into just one verdict, alarm or dismissal, is a distortion of all three. The framework's focus on likelihood, severity, and evidence is what prevents it from becoming a "speculative harm" = "certain harm" or a "proven harm" = "hypothetical harm".

4.3 Governance implications. Answers will vary due to the difference in domains. Mitigation, as in many equity issues, can be defined and mandated for mitigations in harms as well as access and inclusive data. When the effects are unknown but a consequence of the investment, such as in the case of labour, the posture is to prepare and monitor investment in transition support and instrumentation for detecting the effects as they become apparent. As with large scale misinformation, there are harms that can occur in the future and thus there are reasons to pursue anticipatory measures, including measures of provenance and detection, even if some of the harms might be difficult to fix. The main governance recommendation in the framework is to match the strength and timing of the response to the strength of the evidence.

4.4 Relationship to existing taxonomies and regulation. The framework builds on and complements the risk taxonomies and ethics principles that come before it. Taxonomies list things that can go wrong; principles list things that should be valued; emerging regulation specifies things that are required. It is placed between them, providing the evaluative step which brings an enumerated risk and a prioritised response together and is expected to be used in conjunction with institutional tools like algorithmic impact assessments and risk management standards. While this type of process and obligation is provided by those instruments, here the dimensions and matrix serve as a common analytic core that constitutes the substantive judgement these instruments invite.

4.5 Limitations. Conceptually, the framework provides a way to organise and structure assessment, but does not address the empirical questions that it can help to ask, and the evidence on which to apply it is, in some respects, still underdeveloped or very dynamic. The three domains it focuses on are intended, but incomplete: privacy, security, environmental cost and the focusing of economic power are other important elements of the social impact of generative AI that a more comprehensive evaluation would include. The dimensions and matrix provide for structured judgement, but do not eliminate the normative judgements made by assessment about acceptable risk and fair distribution. Any use of the framework should be documented clearly in order to capture these limitations.

5. Conclusion

The social impact of generative Ai is far too significant to be a matter of fear or fascination. In this paper, I have proposed a conceptual framework for shifting from general claims to specific, evidence-based and comparable judgments, specifically in three of the most important and controversial fields: misinformation, labor, and equity. The framework lists the key risks and mechanisms in each domain, assesses any potential risk across the shared dimensions of likelihood, severity, reversibility, distribution, timescale and tractability, prioritises risks in a severity-by-likelihood matrix, anchors the assessment by domain specific indicators, and maps out the different actors who can implement mitigation levers. It reflects the principle of ‘matching the strength of the response to the strength of the evidence' and is about responding to harms, preparing for harms, and monitoring harms, where the harm is demonstrated, and taking anticipatory action, where the harm cannot be reverses easily. When applied in this way, it provides a common language and a standard methodology for researchers, developers, and policymakers to consider social risks from generative AI tools while remaining mindful of their social benefits.

Declarations

Conflict of Interest

The authors have no conflicts of interest to disclose concerning this article.

Funding Statement

This work was not supported by any specific grant from public or commercial funding sources or from not-for-profit institutions or agencies.

Ethical Approval

As no human or animal subject was used in this study, ethical approval was not necessary.

References

[1] R. Bommasani et al., "On the opportunities and risks of foundation models," arXiv:2108.07258, 2021. Available: https://arxiv.org/abs/2108.07258

[2] L. Weidinger et al., "Taxonomy of risks posed by language models," in *Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency*, pp. 214–229, 2022. doi: 10.1145/3531146.3533088.

[3] R. Shelby et al., "Sociotechnical harms of algorithmic systems: Scoping a taxonomy for harm reduction," in *Proceedings of the 2023 AAAI/ACM Conference on AI, Ethics, and Society*, pp. 723–741, 2023. doi: 10.1145/3600211.3604673.

[4] I. Solaiman et al., "Evaluating the social impact of generative AI systems in systems and society," arXiv:2306.05949, 2023. Available: https://arxiv.org/abs/2306.05949

[5] B. Chesney and D. Citron, "Deep fakes: A looming challenge for privacy, democracy, and national security," *California Law Review*, vol. 107, no. 6, pp. 1753–1820, 2019.

[6] J. A. Goldstein, G. Sastry, M. Musser, R. DiResta, M. Gentzel, and K. Sedova, "Generative language models and automated influence operations: Emerging threats and potential mitigations," arXiv:2301.04246, 2023. Available: https://arxiv.org/abs/2301.04246

[7] B. Buchanan, A. Lohn, M. Musser, and K. Sedova, *Truth, Lies, and Automation: How Language Models Could Change Disinformation*. Center for Security and Emerging Technology, 2021.

[8] S. Kreps, R. M. McCain, and M. Brundage, "All the news that's fit to fabricate: AI-generated text as a tool of media misinformation," *Journal of Experimental Political Science*, vol. 9, no. 1, pp. 104–117, 2022. doi: 10.1017/XPS.2020.37.

[9] C. Vaccari and A. Chadwick, "Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news," *Social Media + Society*, vol. 6, no. 1, 2020. doi: 10.1177/2056305120903408.

[10] T. Eloundou, S. Manning, P. Mishkin, and D. Rock, "GPTs are GPTs: An early look at the labor market impact potential of large language models," arXiv:2303.10130, 2023. Available: https://arxiv.org/abs/2303.10130

[11] C. B. Frey and M. A. Osborne, "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change*, vol. 114, pp. 254–280, 2017. doi: 10.1016/j.techfore.2016.08.019.

[12] D. Acemoglu and P. Restrepo, "Automation and new tasks: How technology displaces and reinstates labor," *Journal of Economic Perspectives*, vol. 33, no. 2, pp. 3–30, 2019. doi: 10.1257/jep.33.2.3.

[13] E. Brynjolfsson, D. Li, and L. R. Raymond, "Generative AI at work," NBER Working Paper No. 31161, National Bureau of Economic Research, 2023. doi: 10.3386/w31161.

[14] S. Noy and W. Zhang, "Experimental evidence on the productivity effects of generative artificial intelligence," *Science*, vol. 381, no. 6654, pp. 187–192, 2023. doi: 10.1126/science.adh2586.

[15] E. M. Bender, T. Gebru, A. McMillan-Major, and S. Shmitchell, "On the dangers of stochastic parrots: Can language models be too big?" in *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, pp. 610–623, 2021. doi: 10.1145/3442188.3445922.

[16] J. Buolamwini and T. Gebru, "Gender shades: Intersectional accuracy disparities in commercial gender classification," in *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, PMLR 81, pp. 77–91, 2018.

[17] S. L. Blodgett, S. Barocas, H. Daumé III, and H. Wallach, "Language (technology) is power: A critical survey of 'bias' in NLP," in *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pp. 5454–5476, 2020. doi: 10.18653/v1/2020.acl-main.485.

[18] N. Mehrabi, F. Morstatter, N. Saxena, K. Lerman, and A. Galstyan, "A survey on bias and fairness in machine learning," *ACM Computing Surveys*, vol. 54, no. 6, Art. no. 115, 2021. doi: 10.1145/3457607.

[19] L. Floridi et al., "AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations," *Minds and Machines*, vol. 28, no. 4, pp. 689–707, 2018. doi: 10.1007/s11023-018-9482-5.

[20] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines," *Nature Machine Intelligence*, vol. 1, no. 9, pp. 389–399, 2019.