



Deposited via The University of Sheffield.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/219754/>

Version: Published Version

Article:

Bhalla, N., Brooks, L. and Leach, T. (2024) Ensuring a 'responsible' AI future in India: RRI as an approach for identifying the ethical challenges from an Indian perspective. *AI and Ethics*, 4 (4). pp. 1409-1422. ISSN: 2730-5953

<https://doi.org/10.1007/s43681-023-00370-w>

Reuse

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



Ensuring a ‘Responsible’ AI future in India: RRI as an approach for identifying the ethical challenges from an Indian perspective

Driving Responsible AI in the world’s largest society: RRI as a tool for identifying the ethical challenges of AI in India

Nitika Bhalla¹ · Laurence Brooks² · Tonii Leach¹

Received: 28 April 2023 / Accepted: 23 October 2023 / Published online: 4 December 2023
© The Author(s) 2023

Abstract

Artificial intelligence (AI) can be seen to be at an inflexion point in India, a country which is keen to adopt and exploit new technologies, but needs to carefully consider how they do this. AI is usually deployed with good intentions, to unlock value and create opportunities for the people; however it does not come without its challenges. There are a set of ethical–social issues associated with AI, which include concerns around privacy, data protection, job displacement, historical bias and discrimination. Through a series of focus groups with knowledgeable people embedded in India and its culture, this research explores the ethical–societal changes and challenges that India now faces. Further, it investigates whether the principles and practices of responsible research and innovation (RRI) might provide a framework to help identify and deal with these issues. The results show that the areas in which RRI could offer scope to improve this outlook include education, policy and governance, legislation and regulation, and innovation and industry practices. Some significant challenges described by participants included: the lack of awareness of AI by the public as well as policy makers; India’s access and implementation of Western datasets, resulting in a lack of diversity, exacerbation of existing power asymmetries, increase in social inequality and the creation of bias; the potential replacement of jobs by AI. One option was to look at a hybrid approach, a mix of AI and humans, with expansion and upskilling of the current workforce. In terms of strategy, there seems to be a gap between the rhetoric of the government and what is seen on the ground, and therefore going forward there needs to be a much greater engagement with a wider audience of stakeholders.

Keywords Responsible research and innovation · Responsible AI · India · Focus group · Governance · Ethics · Society

1 Introduction

India spans an area of 3,287,263 sq km in the South Asia region and has a current GDP of ca. US\$3.39 trillion in 2022 [1]. This is the world’s most populous country with a population of ca. 1.42 billion people [2]. India is ranked second in the world in terms of active Internet users. The

Internet penetration rate in India went up to nearly 48.7% in 2022, from just about 4% in 2007. This suggests that almost half of the population had Internet access that year [3]. The emergence of a digital society has also led to a digital divide in India between urban and rural areas affecting digital education and economic opportunities [4]. In addition, it is understood that there is a divide between affluent cities and under-resourced urban areas. The digital divide mainly includes poor digital infrastructure in villages and under-resourced urban areas, limited access to digital facilities, and poor socioeconomic conditions. Other studies have shown that there is an additional second-level digital divide (i.e. those individuals’ skills to use computers and the Internet) between the disadvantaged caste groups and Others [5]. Despite the digital divide, India was ranked #32 in the

✉ Laurence Brooks
l.brooks@sheffield.ac.uk

Nitika Bhalla
nitika.bhalla@dmu.ac.uk

¹ Centre for Computing and Social Responsibility, De Montfort University, Leicester, UK

² Information School, University of Sheffield, Sheffield, UK

global index that shows which countries in the world are best placed to maximise the potential of AI in public service delivery. India has seen the development of a number of initiatives aimed at embedding AI use throughout all levels of government and society [6].

The impact of an emerging technology like artificial intelligence (AI) can be seen as ‘inevitable’ and is already being felt in numerous ways by citizens globally. According to the 2022 IBM Global AI adoption index, 57% of Indian organisations reported having deployed AI, and 27% as exploring AI [7]. Although AI is often deployed with the intention of improving access, quality and efficiency, risks and challenges of AI have also emerged across a number of different areas. Currently, AI is a technology that continues to advance rapidly and the discourse on AI ethics and its governance is also evolving. Globally, an ever increasing number of different sets of ‘AI ethics principles’ have been put forward by organisations, private sectors and various nation states [8, 9].

Governments globally, including India, have started to take a serious stance of incorporating principles and values around AI governance and ethics. Apart from the establishment of principles, however, it is also necessary for India to frame a means of implementing the principles across the public sector, private sector and academia in a manner that balances innovation and governance of potential risks [10]. To note, the strategies which have been developed in India include the National Strategy on AI [11] and Responsible AI [10] (both these strategies are discussed below).

Much of the current discourse on AI in India is government led, and predominantly economically focused, considering mainly the fiscal benefits that this technology offers India [12]. However, there is less discussion of the social and ethical issues generated by the development, deployment and adoption of AI in India presently [13]. These ethical–social issues include concerns around privacy, data protection, job displacement, historical bias and discrimination. The potential for even well-intentioned algorithmic systems to have disproportionate consequences on vulnerable and marginalised communities is also left unconsidered [14].

This research seeks to drive forward the debate on socially beneficial and ethically sound AI in India by considering the current social and ethical risks and challenges as considered by AI experts in India, and utilising a responsible research and innovation (RRI) framework to explore the possibilities for tackling these challenges.

By analysing the results of a number of focus groups on the topic of AI in India, this study aims to explore the unique social and ethical issues related to AI in the Indian context, and considers the possibilities for tackling these challenges through the lens of the RRI AREA framework. As such, this paper seeks to answer the following research questions:

1.1 What societal/ethical challenges related to AI does India face? Can RRI be used as a tool for identifying some of these possible societal/ethical challenges?

This study is part of a series of global AI governance focus groups which aim to gather and analyse insights of how the challenges and opportunities of AI differ across different parts of the world. In particular, India, with its high level of regional influence, its fast-growing and large-scale AI industry, and the introduction of a number of ambitious, cross-cutting government AI initiatives make it a key player in the growing global AI market [14].

This paper begins with an overview of the current state of AI in India, and the possibilities of RRI as a framework for considering social and ethical issues and challenges of AI in India. This provides a basis for the overview of the methodology of the focus groups. The following section reports the findings and insights gained from the focus groups in India. This is used in the discussion section to inform the discourse on current social and ethical issues of AI in India and examine how RRI can be used as a potential framework for analysing and addressing some of these challenges. The paper also offers an overview of the oft-cited case of the Indian COVID-19 app as an exemplar of the types of social and ethical concerns raised and offers comparisons between relevant literature and the findings of the focus groups. The conclusion highlights some of the limitations of this study and considers the next steps to further drive the debate on tackling social and ethical issues of AI in India today.

2 Responsible innovation in AI

2.1 Artificial intelligence (AI)

AI has received particular prominence, as the public interest in the ethics of AI has grown from the mid-2010s onwards [15]. AI may be seen as one of the most disruptive technologies of the twenty-first century, with the potential to transform every aspect of society [16]. While discussing the issue of ethics, the challenges of data privacy as well as exacerbations of bias and racial discrimination need to be acknowledged. The potential for even well-intentioned algorithmic systems to have disproportionate consequences on vulnerable and marginalised communities is also left unconsidered.

The term ‘AI’ is subject to a veritable plethora of interpretations exhibiting a variety of granularity of consideration, which range, for example, from Boden’s assertion

that AI is capable of making ‘computers do the sorts of things that minds can do’ [17] to Bostrom’s concept of a system which possess ‘common sense and an effective ability to learn, reason, and plan to meet complex information-processing challenges across a wide range of natural and abstract domains’ [18]. Given this, we would like to offer some definitions in this area for the purpose of this paper.

- **Algorithms** are a set of ‘encoded procedures for transforming input data into desired output, based on specific calculations’. Procedurally, algorithms run in a series of steps and do not require human oversight or intervention to run [19].
- **AI** refers to digital systems that demonstrate ‘intelligent’ behaviour. AI can be categorised into three main types: artificial superintelligence, which refers to a system that surpasses human intelligence, social skills and knowledge across a broad range of areas¹; artificial general intelligence, in which a system could exhibit intelligence equivalent to, or above, human intelligence; and narrow artificial intelligence, which refers to systems capable of demonstrating human-like intelligence, but only in narrow, specific domains (eg. ChatGPT) [18, 20, 21].
- **Machine learning** refers to the common (and, arguably, the most successful) subset of AI techniques used today, which uses recursive self-improvement to increase task performance over time, and has been available for a number of years, pre-dating the modern concept of AI [14].
- **Data** information, especially facts or numbers, collected to be examined and considered and used to help decision-making, or information in an electronic form that can be stored and used by a computer.

2.2 Artificial intelligence (AI) and India

The development, adoption and promotion of AI have been on the list of priorities of the Indian Government, an approach that rests on the premise that AI has the potential to make lives easier and society more equal [14]. The government over the years has allocated substantial funding towards research, training and skilling in emerging technologies like AI [14, 22]. Furthermore, initiatives like Digital India are aimed at transforming India into a ‘digitally empowered society and knowledge economy’ [23]. This particular initiative attempts to provide a digital infrastructure as a core utility to every citizen, incorporating such digitisation in governance and ultimately leading to the empowerment of citizens [14]. The government has also begun to work towards ensuring that AI technology is made in India,

and made to work for India as well, fitting squarely within its ‘Make In India’ programme, an initiative to promote India as a global manufacturing hub.

Further initiatives such as the AI Task Force and the National Strategy for AI published in June 2018 (#AIFORALL) developed by NITI Aayog (the policy think tank of the Government of India) was an aspirational step toward India’s AI strategy. It covered a host of AI-related issues including privacy, security, ethics, fairness, transparency and accountability. It also focused on reskilling as a response to the potential problem of job loss due the future large-scale adoption of AI in the job market. Over recent years, the growth in the research and innovation of technologies has increased, and India continues to be a key player globally. For example, the deployment of technologies such as AI in sectors such as health, agriculture, education, etc. [22] in India has only grown through the support of state governments, research institutions, leading applications from the private sector and a vibrant evolving AI startup ecosystem [10]. The discourse on AI ethics and governance is also evolving. Globally, a number of different sets of ‘AI ethics principles’ have been put forward by multilateral organisations, private sector entities and various nation states. However, it is also necessary for India to create frameworks that will implement the principles across the public sector, private sector and academia in a manner that balances the promotion of the benefits of AI, with mitigation of the risks specific to the Indian society. The call for *Responsible AI* indicates the establishment of broad ethical principles for design, development and deployment of AI, and hopefully these new ideas will gain currency in India’s policy making.

In 2021, UNESCO member countries adopted the first global agreement on the ethics of AI [24]. The objective of this work was to provide a universal framework of values, principles and actions to guide states in the formulation of their legislation, policies and other instruments regarding AI, in a manner consistent with international law [24]. This recommendation addresses ethical issues related to the domain of AI, the framework of interdependent values, principles and actions that can guide societies in dealing responsibly with the known and unknown impacts of AI technologies and offers them a basis to accept or reject AI technologies. In addition, it considers ethics as a dynamic basis for the normative evaluation and guidance of AI technologies, referring to human dignity, well-being and the prevention of harm as a compass and as rooted in the ethics of science and technology [24]. This recommendation pays specific attention to the broader ethical implications of AI systems in relation to the central domains of UNESCO: Education, Science, Culture, and Communication and Information.

Although international discussions around AI, ethics and social safety seem to be currently premised on Western contexts, India needs to ensure that specific ethical issues

¹ However, it should be noted that the possibility that such a system could ever exist is highly contentious.

that are pertinent to its society are identified and addressed appropriately. One suggested proposal is to use responsible research and innovation (RRI) as an analysis framework to address and mitigate some of these issues and challenges. Although RRI is a heavy topic of debate in Europe among the Science and Technology Studies academic community [25], it can be argued that incorporating RRI initiatives into a country like India, which is rich in cultural and economic diversity, can help to align science and society through engagement with stakeholders and practices. Given the structure of India as largely state driven in some capacities, it is important that any framework used to consider these challenges considers the multiplicity of contexts to be addressed: RRI, with its consideration for value alignment across differing European Nation, offers the scope to consider such a plurality of contexts within India. Promoting Responsible AI is also included within the Indian National Strategy for AI [22].

2.3 Responsible research and innovation (RRI)

This paper focuses on the AREA framework [26], AREA is an acronym for Anticipate (this gives the opportunity to address the possible implications of the proposed research), Reflect (reflection is an opportunity to think critically about current methodology and research practice), Engage (this steps enables the development and embeds a communication strategy for the research) and Act (this looks to develop a strategy on how to act upon the information arising from engaging with stakeholders). Essentially, these four processes or steps help put research in the context of responsible innovation.

The concept of responsible research and innovation (RRI) rose to prominence in 2010, when Schomberg set a vision of this concept and how it could be embedded within policy narratives, taking a largely European policy perspective. The intention to align research and innovation with societal needs and preferences was at the forefront of this vision, and a broad framework was proposed for its implementation under research and innovation schemes around the world [27]. As an awareness of the potential challenges and issues created by newly developed or emerging technologies has grown (alongside an exponential increase in the number of new technologies developed), RRI has become an important framework for considering the implications of these technologies at an early stage in the development process [28]. Of the range of currently emerging technologies, artificial intelligence (AI) is often the most divisive in both consideration of its nature and expectations for its impact on both the individual and society.

The emphasis and origin of RRI have largely been in the Global North, whereas reference to what RRI or RRI-like practices mean in the context of the Global South is limited

[25]. For example, the cultural differences between Europe and India govern how they interact with other societies and how they engage with RRI research. Therefore, using RRI in India will need to be implemented in a way that is sensitive and can be useful and have a positive impact on its citizens.

2.4 RRI in India

The practice and implementation of RRI is still a developing concept in India and not part of the official discourse on science and technology, although many elements of RRI, including gender and science education, are present in various policies and programmes in different forms in India. Literature suggests that science policy in India is primarily concerned with science and technology for societal and national development [29]. Science policy has been sensitive to the changing dynamics in global S&T and to the need to harness emerging technologies. The RRI pillars of Ethics, Gender, Open Access, Science Education and Public Engagement, are present in the science policy in a variety of ways.² In recent years, ethics in policy and practice has gained prominence in life, bio-, and health sciences, and guidelines and rules have been updated, revised and rewritten, to make the most of the practices compatible with best practices from around the world [29]. Gender and participation by women has gained the attention of policy makers, scientific academics and academic institutions, resulting in specific policies and programmes, but the lack of a comprehensive framework to tackle these issues coupled with institutional, social and cultural factors act as a barrier to improving equality in this area. Open aAccess as an initiative is supported by a wide range of policies and initiatives, but lacks uniform policies and principles, along with the failure of institutions to work together to connect silos and make open access more meaningful and truly accessible. Science education in the Indian context differs from the EC definition as outlined in RRI. Still, science education gets wide-ranging support in India from the government, although much remains to be done. Public engagement is not explicitly supported in policy in India and is often treated as an equivalent of science communication. This is premised on the belief that there is a deficit in public understanding of science which should be addressed more by educating the people, communicating to them and enhancing their understanding and appreciation for science. Thus monologue is preferred over dialogue or engagement, although some initiatives have tried to reverse this [29].

The current attitudes around RRI in India require bridging the gap between normative acceptance and practice. India as a country is large and complex and relies on individual state-driven government policies. RRI in theory and

² The terminology used may vary, but the principles are the same.

practice can benefit from interaction with ideas and practices developed in India such as access, equity and inclusion, scientific temper and scientific social responsibility. However, there are some limitations of adopting the RRI approach in a Global South context that need to be considered; for example, RRI has been tested and implemented more in Europe than elsewhere. Therefore, for RRI to be relevant in India, the approach must be contextualised and that contextualisation has to be sensitive to issues in STI and the societal needs (Srinivas, 2022). Compared to Europe, the Global South, is less techno-economic and more community oriented [25]. Therefore for RRI policies to be widely accepted in India, there must be a dialogue between policy makers and the public. The step to becoming more responsible must be initiated on two fronts: in government policies, and in the organisational framework and people's mind set. It is important to understand the relevant characteristics of RRI and RRI-like practices according to the cultural and socio-political context, and how social drivers and barriers stimulate and hinder RRI practice in India.

Although AI in India is often deployed with intentions of improving access, quality and higher efficiency, risks and challenges of leveraging AI have also emerged across a number of different areas. If these risks are not managed responsibly, it is likely to have a significant negative social and economic impact. One possibility of identifying and addressing these issues from the outset is using RRI as an analysis tool.

3 Focus group methodology

This research study was focused on the challenges of AI at the societal level. Therefore, to investigate these challenges clearly from the ground up, a focus group methodology was adopted. Focus groups are an 'informal discussion among selected individuals about specific topics' ([30], p. 73), designed to explore people's 'attitudes, opinions, knowledge or beliefs' ([31], p. 185) and 'enable the development of collective understandings of shared problems—and (often) solutions to these problems.' ([31], p. 186).

As the research questions in this case move beyond the remit of simply the individual perspective, the deployment of the focus group in this case allows for the consideration of a socially constructed understanding of these issues highlighted by the interactions between participants. Given that the topic of AI is such a polarising one, giving participants the opportunity to engage with others to explore a range of collective responses to the research questions allows for an understanding as to how and why such perspectives develop [32]—not simply which views are most prevalent.

Whilst a focus group size usually consists of between 6 and 12 participants, given the specific knowledge, expertise

and experiences of the participants, it was deemed appropriate to work with smaller focus group sizes of between three and five participants in an approach similar to Kruegar's [33] 'mini' focus groups [34–36]. This gave each participant the space (and/or time) to express their own, considered and relatively expert view on the subject whilst also leaving room for discourse and debate between the participants.

Focus groups were conducted online via the Zoom meeting platform. This decision was pragmatic, given the severity of travel restrictions enforced due to the COVID-19 pandemic at the time the focus groups took place.

3.1 Participants

Participants were recruited via a targeted email campaign and registered their interest in taking part in the focus groups via an Eventbrite registration web page, where they were also screened for suitability for the focus group. To ensure that the focus group participants had a suitable depth of knowledge for both the topic of the focus groups (AI) and the context in which the researchers were interested (the contemporary Indian context), eligible participants were required to meet the following criteria: over the age of 18 years; an AI professional or expert who has a considerable degree of knowledge in this field; knowledgeable about the potential impacts of AI in the focus group country (India); and familiar with the current situation and context of the focus group country (e.g. a national that currently lives and works there, or has been away from the country for less than 5 years, or a non-national who has lived and worked there for at least the last 5 years). This ensured that the possibility of participants with only a minimal knowledge of the current state of the art in AI technology, as well as those that have an outdated and/or limited knowledge of AI development, deployment and adoption of AI technologies *in India specifically* were screened out. This gave the research the best possible chance of reflecting current, timely, knowledgeable and relevant views on the societal challenges of AI in India and the concomitant opportunities of RRI within this scope. All participants who were eligible according to these criteria were provided with detailed information about the focus groups and gave informed consent to participate.

Four focus groups were conducted between November 2020 and February 2021.³ A total of 13 participants took part, with between 3 and 4 participants in each focus group. Participants were able to select which focus group they wished to participate in to allow the maximum number of participants to take part. This allowed for interactions to develop between a heterogeneous group of participants in

³ As this study was conducted prior to the launch of ChatGPT in November 2022, the study did not include reference to that and related large language models.

each focus group, with the aim of encouraging diverse dialogues to develop what reflected the background and specific expertise of each participant.

Each focus group was facilitated by either one or two members of the AI Global Governance Research Group based at De Montfort University, UK, who are knowledgeable in this area of research. Because a number of focus groups were conducted, the facilitators adopted a standardised presentation to guide the focus group to ensure that each focus group was exposed to identical information about the study and the framing of the questions. This presentation included a recap of the information to participants provided at the point of registration, an explanation of the procedure of the focus group, data management (including the recording process) and the questions that the focus group participants were asked. This reduced the risk of researcher bias and ensured that all participants were at roughly the same informational place from which to begin the focus group discourse.

After providing informed consent, all participants were provided with the link to attend the focus groups. Each focus group lasted between 43 and 80 min in duration, depending on the number of participants in the group and the intensity of the dialogue, with a mean average focus group time of 57 min. The discussions were audio recorded and transcribed verbatim. Prior to the commencement of the recording, the participants were asked to introduce themselves, in order that the participants could get to know each other with the intention of facilitating discussion and avoiding the possibility of a sort of group interview, where each person simply provides their own opinion in relation to each question, without interacting with the other participants or their answers, which would limit the usefulness of the focus group approach.

The facilitators adopted a semi-structured approach to the focus groups, with six pre-determined questions asked of the participants, with follow-up questions asked as and when the facilitator felt that more information could be pertinent and to encourage group discussion. All participants were encouraged to engage with each question, although a response to every question was not required from each participant.

The six pre-determined questions were developed by the researchers, which were designed to explore potential societal challenges and benefits of AI in the specific country. While many views on the nature of, as well as the development, deployment and adoption of, AI technologies are polarised, it was not considered reflective of the range of views on this topic to focus the questions solely around the potential challenges; as such, questions on both the challenges and benefits of AI were included to encourage a plurality of views to be expressed and considered and to stimulate debate across the breadth of this topic. The following questions were provided to all participants with the aim of

generating a more holistic overview of the considerations of AI technology that examines not only the societal, but also the cultural and political conditions within which such considerations are necessarily situated:

1. What are the most important possibilities and concerns about the use of AI within the next 5–15 years in [the focus group] country?
2. In what ways are current social, cultural and political conditions in [the focus group] country likely to influence how AI is or will be used?
3. Who do you think will benefit most from AI in [the focus group] country, and who do you think will be disadvantaged and benefit the least? And why do you think these differences exist?
4. In which ways do you think that global inequalities will affect the adoption, development and regulation of AI in [the focus group] country?
5. In previous questions you identified a range of potential challenges and benefits for the use of AI in [the focus group] country. How can these challenges be addressed? How can benefits be assured?
6. What do you think will be the barriers to achieving the possible solutions discussed in the previous question?

The focus group manuscript was professionally transcribed, and the transcripts analysed using NVivo Pro 12. A combined approach to coding was adopted, as it was anticipated that the global nature of the focus groups would require a more flexible approach. A preliminary round of data analysis was conducted by a core group of researchers on a sample of transcripts to determine an overarching framework for coding based on the key, identifiable, themes. This broad framework of thematic nodes and subnodes was designed to be added to and adapted as each additional focus group transcript was coded. The responses were considered within the context of an RRI framework—that is, the extent to which an RRI approach was potentially lacking in relation to identified challenges, and the extent to which there were possibilities to address these challenges through adopting an RRI approach.

3.2 Respondent demographic

To comply with our pseudonymisation policy, we needed to prepare the transcripts for the focus groups in the following way, please see Table 1 below which represents the code names given to participants. Person labels such as “speaker 1”, “speaker 2”, “speaker 3”, etc. as well as participants’ names in the transcripts need to be replaced with standardised code names.

The code names were constructed as follows:

Table 1 Respondent number and demographic of participants

Respondent number	Demographic
R1	FG1_P2_[PRO-H&SS-f]
R2	FG1_P1_[PRO-H&SS-f]
R3	FG1_P3_[PRO-H&SS-m]
R4	FG4_P1_[PS/C-m]
R5	FG4_P2_[PRO-H&SS-f]
R6	FG2_P2_[PRO-FS-m]
R7	FG2_P1_[PRO-T&E-m]
R8	FG2_P3_[PRO-FS-m]
R9	FG3_P2_[PS/C-T&E-m]
R10	FG3_P1_[PRO-H&SS-f]
R11	FG3_P3_[PRO-T&E-m]
R12	FG4_P3_[PRO-T&E-m]
R13	FG3_P3_[PRO-H&SS-m]

focus group number (e.g. FG3) + participant number (e.g. P1) + [abbreviations of person descriptors in the following order: type of organisation + disciplinary background (for PROs only) + gender].

3.3 For type of organisation

- Private sector/company (PS/C).
- Public research organisation (PRO) (e.g. universities and publicly funded research institutes).
- Governmental body (GB).
- Civil societal organisations (CSO).

3.4 Disciplinary background

- Formal sciences (FS) (e.g. maths, statistics and theoretical computer science).

- Technology and engineering (T&E) (all technical and engineering-related disciplines that work towards practical applications, including applied computer science).
- Natural sciences (NS) (e.g. physics, chemistry, life sciences).
- Humanities and social sciences (H&SS) (including law and economics).

3.5 Gender

- Female (f)
- Male (m)
- Gender not disclosed (gnd)

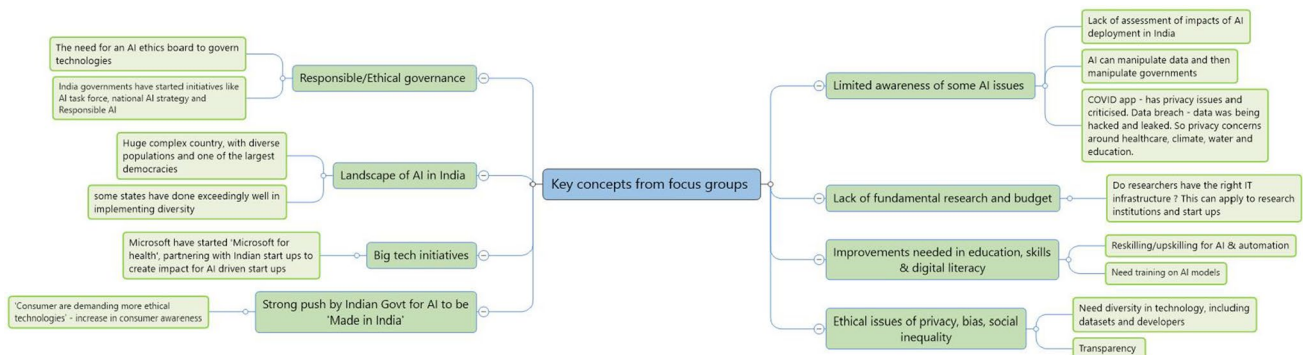
4 Results

The nature of focus groups allows for discourse to occur between participants that can move in organic directions (even allowing for the performativity required by the artificial nature of focus groups themselves [37]). Therefore, the findings of the focus groups presented here reflect responses that have a clear relevance to the topics of societal challenges posed by AI in India and the possibilities for RRI to mitigate some of these challenges. This is in no way to suggest that participant responses on a broader range of topics do not hold intrinsic value—but simply that they are beyond the scope of this paper. Figure 1 represents the key concepts discussed by the participants.

4.1 Bias and privacy

Firstly, participants were asked about near term concerns [38] about AI, specifically in India. The focus groups described the following key concerns.

Firstly, ‘the biases that [AI] brings’. This is one issue that resonated with most participants, for example datasets are created in the West by ‘predominantly white, middle class

**Fig. 1** Key conceptual findings from the India focus group study

males’ (R1) and then used by the diverse Indian population. One such suggestion proposed by a participant was to have ‘a diverse workforce which can hopefully be one way of mitigating bias’. However it was highlighted that ‘the concern is not just having representative stakeholders in the actual coding of design systems’ (R1), as there may be ‘implicit bias that [the coders] just do not see.’ A significant cause of concern was around privacy during the launch of the COVID-19 app. The data of patients was being ‘hacked and leaked’ (R2), and the misuse of data concerns many sectors in India including healthcare, climate and education. Another key issue is that India does not have a data protection bill nor does it have an adequate data privacy legal framework. Therefore, how data is generated, accessed, used and stored is not clear or transparent. Participants were not clear about ‘How are these decisions [are] being made?’ and ‘was it with the consent of the people?’ (R2). One participant mentioned that the government is currently trying to ‘apply India’s constitutional framework to emerging technologies and to AI as well’ (R3) which will hopefully address some of these issues.

Secondly, participants were asked to reflect how AI is or will be used in India given its current social, cultural and political condition. This question was approached in a number of ways by the focus group participants, although the political landscape featured as a primary source of discussion, with participants highlighting political ‘misinformation’ (R4) campaigns and ongoing bureaucracy as factors affecting the development and use of AI in India. Participants also raised concerns about the lack of existing legislation in India to protect the privacy of individuals within the context of AI, arguing that a lack of data protection law leaves individuals in India without a say in how and in what ways their data is used. External actors were also considered in the responses to this question, particularly Western ‘big tech’ companies, who were framed as exploiting legislative loopholes to gain access to large personal datasets irrespective of the people’s choice—furthermore, the exploitative nature of Western big tech companies in relation to India was specifically characterised as a colonisation of technology, suggesting these practices are considered an embodiment of a more systemic, unilateral and top-down approach to AI that fails to centre on the needs of Indian citizens.

4.2 Social inequality and power symmetries

In terms of thinking about the impact around the inequalities that AI brings to society, the general consensus of participants was that technologies, including AI, are essentially a reflection of existing power asymmetry in society. As it is a specific set of people, most of whom are privileged enough to have access to it, this also includes the people who are largely coding in India, i.e. ‘Mostly men from upper class

backgrounds’ (R1). One participant went on to explain that ‘if the policy is structured in a way that allows large technology companies based in the US to get away with competition law violations, and to use the data of citizens without necessarily having an adequate framework for exploitation, then it can exacerbate existing power asymmetries’ (R1). The participant alluded to the viewpoint that ‘it was largely the private sector and the corporates that benefit, often at the expense of the individuals that are there’ (R1). It was felt that those who would benefit the most from AI are ‘always’ the big tech companies who have access to more data, greater IP rights and have the potential to monopolise their market for greater profit. In another example, the topic of facial recognition technology came up with the example of the government using the technology to conduct a type of ‘unconstitutional surveillance’ (R1) before the pandemic. The participants mentioned that there is limited consultation with ‘diverse stakeholders who have come from diverse socio-economic backgrounds and have lived diverse realities’ (R2). These societies will be the most disadvantaged. Furthermore, the issue of employment and the replacement of jobs by automation and robotics was a significant one that resonated amongst the participants. To mitigate these risks, participants thought it is important to ensure there is an ‘upskill’ such that new job opportunities can be created and people are not easily replaced by AI methodologies. However, from the education perspective AI was largely seen as a benefit, such that students from the most rural parts of India can potentially have access to education and teaching, which may have not been possible otherwise. Participants also alluded to benefits such as AI being used as a ‘teaching assistant model that could be used for class management and creating a better attendance ecosystem.’ (R5) It was suggested that AI could help towards teaching and create more value for the students. Other sectors that participants felt could benefit from AI are the manufacturing industry, finance, industrial Internet of things and healthcare.

Participants were then asked to consider the impact of globalisation on AI in India. Many of the responses to this question broadly considered aspects of technological colonialism. In response to this question, a number of participants highlighted issues around language as a source of potential inequalities—specifically that the default language for AI systems (the majority of which are developed by Western tech companies) is English, and that most AI applications adopt an ‘English first’ approach. Given the diversity of languages and dialects spoken across India (anything from 22 to 700), this status quo is considered to effectively exclude groups of potential users from accessing AI technologies based on a lack of English language skills. Furthermore, the influence of big Western tech companies was seen as shaping ‘technological [...] [and] regulatory standards via an adherence to the existing regulations of the companies’

base nation rather than developing or adopting technological and regulatory standards more closely aligned with the Indian context. However, some participants more positively reflected on the role of Western big tech influencing AI in India, describing it as a system of ‘collaboration with the West’ whereby the investment of big Western tech companies is seen as ‘capital [...] finding its way into Indian markets’ (R6).

4.3 Mitigation strategies

To try and mitigate some of the challenges of certain discrimination (in society), data privacy and freedom of speech, one approach mentioned was through regulation or technological solutions. Furthermore, raising ‘awareness of the developer/technologist’ and whether they understand the problems that might occur as a result of their innovation. This relates to why the concept of ‘Responsible AI’ is important, as it could help with this context. It was also highlighted that having a ‘human in the loop’ (R2) would help to address some concerns around algorithmic transparency, such that there is not full reliance on the decision made by the algorithm. The idea of AI explainability is important, but for this to be effective there needs to be some level of competency (i.e. through education and training) to understand and override these systems, if necessary. One response was that ‘trying to go for a hybrid approach in India rather than complete an AI approach’ (R7), would be beneficial. Overall participants suggested that having effective ethical governance of AI was important. India is the services super powerhouse of the world so the implementation of regulation and the presence of an Ethics board would only help its economy and country flourish.

After asking about strategies to promote the benefits and address the risks of AI in India, participants were then asked if they were able to identify any barriers that might prevent or obstruct these strategies. Many of the answers to this question could be characterised as considering the exceptionally disruptive nature of AI as unique in creating barriers to solutions to the above mentioned challenges by the intrinsic speed of development and adoption of these systems—that there is nothing unique in and of the challenges posed by AI, beyond that the speed of development is at a ‘much much faster pace [...] than with other technologies earlier’. Therefore, there is little time to adequately react to challenges and implement solutions. Some responses to this question focused on a lack of suitably skilled workers [and corporations that employ such a necessarily high level of skill] that would be capable of dealing with some of the more intransigent technical problems outlined above, and the possible impact on confidence in AI systems based on this. From a more general requirement for reskilling workers in relation to automation and job displacement, a lack

of funding for upskilling and retraining workers was considered a barrier to maintaining high levels of employment in a post-AI, post-low-skilled environment. There was also a great deal of mistrust expressed in relation to there being adequate ‘political will’ to ensure that appropriate technical and regulatory safeguards are put in place to protect citizens, with one participant expressing concern that the ‘market standpoint’ (R2) would override any ‘moral standpoint’ in this respect. Beyond this, there were concerns that any regulation or standards put in place might not be followed or adhered to. In addition, the breakdown between knowledge of AI systems and their areas of application was highlighted, for example where healthcare related AI systems are in place, a lack of knowledge about these systems on the part of healthcare workers and other relevant non-technical actors can act as a barrier to adoption—and in this specific case, a lack of knowledge on the part of the practitioner can dissuade patients from adopting or sharing data with such a system. From a more practical perspective, a lack of investment in the right infrastructure was seen as a key barrier to developing a budding startup ecosystem where access to the best resources is made available to both industry as well as research universities. Figure 2 below provides a summary of the ethical AI challenges identified by the participants.

5 Discussion

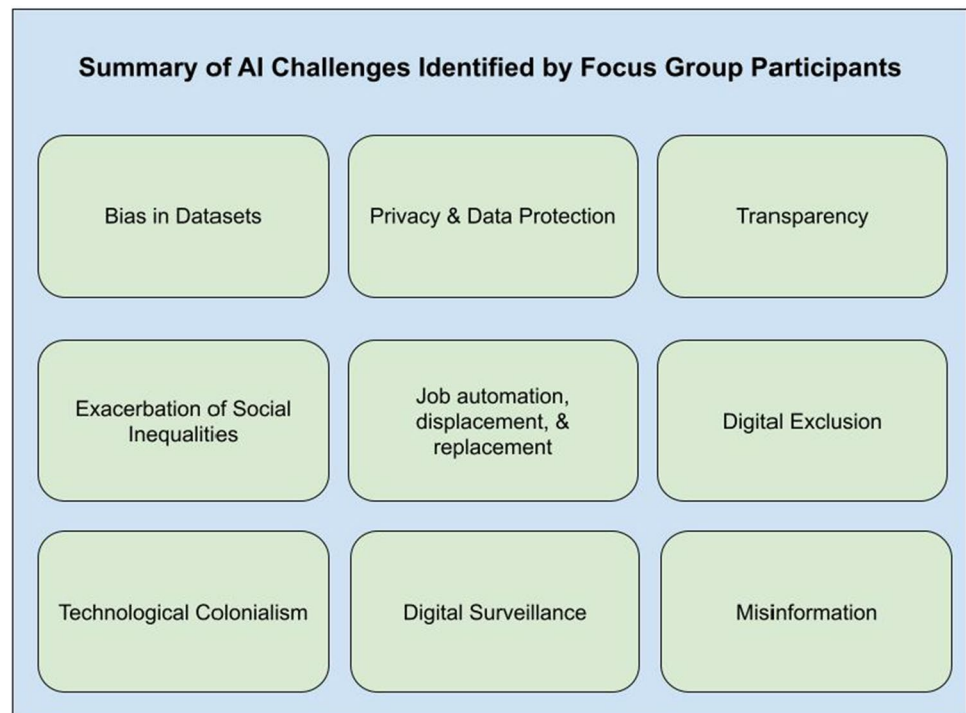
5.1 Embedding RRI into the research and innovation process in India (AREA)

The participants of the focus group have outlined a range of social and ethical issues related to AI that they see as impacting India specifically. RRI not only offers a framework from which to identify and consider these challenges, but also offers scope to be utilised in the development of ways to address these challenges. Here, we will utilise the Anticipation, Reflection, Engagement, Action (AREA) framework to explore the ways in which RRI could be used to address the issues raised.

5.2 Anticipation

In the first instance, many of the focus group participants discussed the challenges of AI in India in terms of deficits; deficits in knowledge, funding, training and infrastructure. The lack of fundamental research and IT infrastructure suggests that these are in need of improvement, together with available funding which is limited for AI researchers, either in educational institutes or startup companies. Additionally, there is a lack of awareness of AI in general and the issues which can arise from the algorithmic system. This is a challenge among the public as well as policy makers,

Fig. 2 Overview of key AI challenges identified by focus group participants



with common misconceptions about the nature, potential and applicability of AI being seen as prevalent across both groups. If this lack of capacity, coupled with previous flaws in homegrown AI technology solutions, continues without intervention, then technological colonialism is likely to result in India being exploited by large Western multinational corporations—in addition to being left behind in technical innovation and the inability to protect its citizens from exploitation.

Privacy and data protection, as well as associated data breaches came up in the focus groups a number of times. In particular the Indian COVID-19 app was used by over a third of participants as an exemplar of the impact of issues in this area. Similar issues were identified in applications used in other fields such as climate, water and education. One notable factor in discussions about privacy and data protection concerned the current lack of legislation in this area, with some participants suggesting that an Indian version of GDPR, to legislate for the protection of people's personal data, might go a long way to building trust in the processes around data collection, and sharing in the context of AI systems developed and deployed in India.

5.3 Reflection

One of the key issues identified in relation to the AI systems currently deployed in India is that it is significantly relying on datasets from other parts of the world (most commonly from the West). This results in India becoming a 'test bed'

for AI rather than a developer (R2), and the (over) reliance on imported datasets is seen to be having a direct impact on the quality of outcomes seen in these systems in India. In particular, discriminatory outcomes are considered a common, negative side effect of systems being trained on a dataset comprising non-local populations, and the lack of diversity in datasets will have implications for India's diverse society and exacerbate issues concerning bias in AI systems and racial discrimination, because the dataset is unlikely to be representative of the population in India.

Improving the quality of AI education with a focus on 'Responsible' research, and upskilling the current AI workforce were suggested as means to improving the quality, safety and security of AI systems developed under the homegrown principle. However, this solution requires an extended time frame to see the benefits of improved education within the workforce and the knock-on impact on the quality of the systems being developed.

One of the most commonly identified ethical issues around AI was a lack of transparency; not only in terms of opacity in the manner the system operates or takes/makes decisions (commonly referred to as a 'black box' system), but also in terms of the impact that the system will have for the people of India. This was considered across two vectors: firstly, systems being deployed with no assessment of the impact on citizens which was discussed primarily in relation to automation and job displacement (which will also be discussed later in this section), and secondly, in terms of the risks posed by AI systems to

fundamental rights. In particular, in the latter case, political autonomy and democratic rights were considered to be heavily impacted by opacity in relation to AI systems. Where AI systems have been deployed for political purposes (for example, by the government), a great deal of suspicion was expressed regarding the possible benefits of the system for the citizens of India balanced against risks posed to democratic rights and freedoms. The government needs to collaborate with relevant stakeholders [39], such as AI practitioners and the public, to evolve standards and guidelines to ensure that AI technology remains socially beneficial while contributing to the economic growth of the country [39].

Concerns about the impact of automation and job displacement as a result of the deployment of AI technologies is a concern that features prominently in societies around the world [40, 41]. This is equally true in India, where the high proportion of un- and low skilled workers are considered to be at high risk of unemployment via automation. One participant mentioned ‘the inequality we have today in society would increase due to AI, and the whole income gap would increase due to the use of this technology if not done in a welfare state of mind’ (R9). In addition to this, India is felt to be particularly vulnerable to this in the shorter term, given a perception that India is being used as a ‘test bed’ for Western AI technologies that increase the likelihood of technologies leading to automation and job displacement being deployed at an early stage. Education is promoted as a key solution to the risks posed by AI in relation to automation and job displacement, with a need to upskill the workforce being seen as the most sustainable way to address this issue. However, systemic issues within the education system are seen as a barrier to successfully upskilling the workforce in a timely manner. The governance of India is largely state driven, and this devolvement of education policy and provision at the state level has created the perception of a ‘postcode lottery’ in which the chance to upskill or retrain in the face of automation is accessible only to those in the states which have invested in developing such training programmes. With no provision at the national level, this could exacerbate inequalities between communities/states within India. One of the possibilities mentioned to mitigate the risk of high rates of automation and resultant job losses is to focus primarily on ‘hybrid’ or human in the loop systems, where the AI system requires ongoing human oversight in the course of its use. This would both ensure continued human oversight and that a proportion of automated jobs could be absorbed back into the industry.

5.4 Engagement

Given the breadth and complexity of the challenges faced in developing and deploying trustworthy, responsible, and ethical AI in India, many focus group participants suggested that more budget needs to be spent on infrastructure as well as research and education to begin to address these challenges (R9). This suggests a need for Responsible Research & Innovation in the educational sector which can address some of the challenges India faces, and going beyond this, there is a need for research to be conducted in a ‘responsible’ manner via ‘allocation of budget to the research and education’ sector (R10), and more broadly to address societal concerns, and a framework such as RRI can be used to tackle some of these challenges.

Given a number of high profile events that have highlighted some of the ethical issues associated with AI technologies, it is unsurprising that the focus groups identified a trend towards the public wanting, and to some extent, expecting, the development of more ethical AI technologies. However, the best approach to ensure this is hotly debated, with a number of approaches being mooted. Firstly, there is a call for AI for social good, to ensure that systems being developed should tackle some form of societal issue. Secondly, there is a need identified for robust AI governance in India, which is felt as lacking in part due to a level of suspicion of governmental practices in relation to bureaucracy and corruption. There is a ‘lack of privacy policies that would allow a more ethical way of sharing data’ (R11). There is a call for the creation of an independent AI ethics board to address concerns around the development, and use of AI in India by multinational corporations, homegrown industry developers and government developers.

Furthermore, it was suggested that the strategy document did not sufficiently explore societal, cultural and sectoral challenges to the adoption of AI. The reflections from the participants also suggested that the general public is calling for more ‘responsible’ AI as the consumer awareness of ethics has increased. A more recent publication on Responsible AI (approach document for India) was published in February 2021 [10]. This highlighted that there is a potential for large-scale adoption of AI which can boost the country’s annual growth rate by 1.3% by 2035. The paper tries to establish broad ethics principles for design, development and deployment of AI in India and discussions have started on the plan to introduce ethics of AI in the mainstream university curriculum to encourage the youth to explore responsible and unbiased AI systems in the near future. Other benefits which India is taking advantage of is that students in the most rural areas are now receiving education due to the availability of AI technology. This initiative allows for inclusivity and aims to reduce the amount of social inequality. India is becoming a ‘centre point of hub of AI adoption’ (R12) and

working with countries such as Israel and Africa in sharing data. Companies such as Microsoft also see the potential in India to be a key global player for tech innovation, for example ‘Microsoft for Startups’ is a new digital and inclusive platform for startup companies in India, offering over US\$300,000 worth of benefits and credits, giving startups free access to the technology, tools, and resources they need to build and run their business.

6 Conclusion and potential actions areas

A number of key themes and narratives emerged from these focus groups which highlighted the need for RRI practices to be incorporated across a number of vectors to support India in promoting the benefits and mitigating the risks of AI in such a diverse and unique society. The areas in which RRI could offer scope to improve this outlook include education, policy and governance, legislation and regulation, and innovation and industry practices. RRI offers the scope to identify and consider these challenges, but also the AREA framework can be used to support the development of ways to address these challenges. In this paper we adopted the Anticipation, Reflection, Engagement, Action (AREA) framework to explore the ways in which RRI could be used to address the social and ethical issues and challenges raised by the focus group participants.

Some significant challenges described by participants included the lack of awareness of AI from the public as well as policy makers. India’s access and implementation of Western datasets can result in; a lack of diversity, exacerbation of existing power asymmetries, increase in social inequality and the creation of bias. The potential replacement of jobs by AI was also a concern, and participants explained that they would not be in favour of complete automation, but a hybrid approach with expanding and upskill of the current workforce.

The case of the COVID-19 app was extolled as a prime example of a home grown application that (at the time) lacked transparency and explainability, and this impacted the trustworthiness of the system. The app has been criticised in literature and by a third of the total participants. The presence of privacy and data breaches suggests that India would benefit from more investment into education relating to the safety and trustworthiness of AI systems and the underpinning infrastructures. This is a key reason why specifically including an RRI dimension in education is important to ensure the next generation of developers have the awareness and ability to engage with these issues and their likely impacts.

The participants also alluded to strategies developed by the government around Responsible AI and the creation of an AI task force. Respondents recognised that India

was trying to collaborate with big technology firms, such as Microsoft, in partnering with India startup companies, to try to foster positive AI impact through the inclusion of experienced innovators. Furthermore there has been a push from the Indian Government for AI to be ‘made’ in India, which requires an understanding of both the technical and the social and ethical dimensions of the technology to be successful.

Other aspirations and recommendations from the focus group participants included; better collaboration amongst researchers, tech companies, and government; greater investment towards IT infrastructure that would harness ‘Responsible’ research; and the need for the implementation of an ethics board to ensure effective governance of AI technologies moving forward.

As we conclude this paper, we would like to highlight some comparisons between literature and the participant perspectives. For example, AI task force and National strategy for AI suggest that both government and public are showing awareness for ethics. This was also reflected in the focus groups where participants suggested an ethics board to effectively govern AI technologies. Although the Indian government has a series of initiatives and set aside funding to increase the digital infrastructure in India. Some participants aired that IT infrastructure still needs to be improved and there needs to be ‘better collaboration,’ as current initiatives seem to be carried out in a top-down approach. One way to enhance collaboration is via stakeholder engagement and consultation with the public and civil society organisations. Again, using RRI as a framework could provide the tools to address some of these issues addressed here. Furthermore, despite the government’s ‘Make In India’ programme, participants still felt that India is used as a ‘test bed’ for AI from the West. For this shift to occur the discourse of Responsible AI and ethics would need to be embedded into the educational system and digital culture of India. As literature suggests RRI is still an evolving concept in India, and this was not mentioned specifically by any participant during the focus groups.

One of key points from the literature is that government initiatives are fiscally-led, i.e. the funding is generally aimed at the technology industry for incentives of innovation and economic development. However, the focus group participants offer a very different perspective on the impact of these initiatives on the ground, with a lack of investment and infrastructure being seen as continuing barriers to the development of a Responsible AI ecosystem in India that will benefit the citizens. Understanding the disconnect between policy and the public’s experience of the AI ecosystem will be vital in ensuring that future initiatives are able to bring the public on board—a greater drive for stakeholder engagement will be essential in shifting this paradigm in a mutually beneficial direction. Given that there is demonstrable awareness

of current AI strategies within the focus group participants, there is a clear opportunity to move beyond the ‘top-down’ model of policy development and engage stakeholders across the country in consulting on future strategic directions for the AI industry that align with the needs and expectations of the public as well as industry and government.

Given that this research highlights a dislocation between policy and the perceived social and ethical issues around AI in India today. This also highlights the opportunities to engage with a broad array of stakeholders to develop AI systems and grow an AI industry that is beneficial to all. The findings of this research is likely to be of interest to policymakers, those developing AI technology in India, and academics working in the broad areas of AI and AI ethics in India.

6.1 Limitations of this study and future directions

Given that a mini focus group approach was adopted for this research, the inherent issues of a small sample size are relevant to this research and as such the views expressed do not necessarily represent the whole Indian population. Future research could seek to expand on this by utilising a larger sample. This is also true of the choice to focus this research on a sample with existing knowledge and experience of AI and its associated issues. Whilst this approach has allowed for the development of a broad overview of current social and ethical issues of AI in India, this does offer primacy to academic and industry schools of thought. Future studies could consider a wider range of stakeholders to ensure a plurality of experiences and views are represented. Finally, the employment of a voluntary sampling process means that the sample used in this study is skewed towards those already knowledgeable and interested in the topic of social and ethical issues related to AI. A more robust sampling method could be used in any follow-up research to ensure a broad spectrum of views are reflected in the results.

Declarations

Conflict of interest The authors declare no conflicting interests.

Ethical approval This focus group study has been approved via the The Research Ethics Committee at De Montfort University, Leicester, UK.

Consent for publication All authors have reviewed the manuscript and have given consent for publication.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are

included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

1. IMF. India: Gross domestic product (GDP) in current prices from 1987 to 2028 (in billion U.S. dollars). Statista, 2023. Accessed: Oct. 17, 2023. <https://www.statista.com/statistics/263771/gross-domestic-product-gdp-in-india/>
2. Hertog, S., Gerland, P., and Wilmoth, J.: India overtakes China as the world’s most populous country’, United Nations, 153, 2023. <https://www.un.org/development/desa/dpad/publication/un-desa-policy-brief-no-153-india-overtakes-china-as-the-worlds-most-populous-country/>
3. Hootsuite and We Are Social. Internet penetration rate in India from 2007 to 2022. Statista, 2023. <https://www.statista.com/statistics/792074/india-internet-penetration-rate/>
4. Laskar, M.H.: Examining the emergence of digital society and the digital divide in India: a comparative evaluation between urban and rural areas. *Front. Sociol.* **8**, 1145221 (2023). <https://doi.org/10.3389/fsoc.2023.1145221>
5. Rajam, V., Reddy, A.B., Banerjee, S.: Explaining caste-based digital divide in India. *Telemat. Inform.* **65**, 101719 (2021). <https://doi.org/10.1016/j.tele.2021.101719>
6. ‘India AI Home’, India AI. Accessed: Oct. 17, 2023. <https://indiaai.gov.in/>
7. ‘IBM Global AI Adoption Index 2022’, IBM. 2022. <https://www.ibm.com/downloads/cas/GVAGA3JP>
8. Ulnicane, I., Knight, W., Leach, T., Stahl, B.C., Wan-jiku, W.-G.: Framing governance for a contested emerging technology: insights from AI policy. *Policy Soc.* **40**(2), 158–177 (2021). <https://doi.org/10.1080/14494035.2020.1855800>
9. Munn, L.: The uselessness of AI ethics. *AI Ethics* (2022). <https://doi.org/10.1007/s43681-022-00209-w>
10. NITI Aayog. Responsible AI #AIFORALL. 2021. Accessed: Mar. 23, 2023. <https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf>
11. NITI Aayog. National Strategy for Artificial Intelligence. 2019.
12. Cath, C.: Governing artificial intelligence: ethical, legal and technical opportunities and challenges. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **376**(2133), 20180080 (2018). <https://doi.org/10.1098/rsta.2018.0080>
13. Sethi, A., Tangri, T., Puri, D., Singh, A., Agrawal, K.: Knowledge management and ethical vulnerability in AI. *AI Ethics* (2022). <https://doi.org/10.1007/s43681-022-00164-6>
14. Marda, V.: Artificial intelligence policy in India: a framework for engaging the limits of data-driven decision-making. *Philos. Trans. R. Soc. Math. Phys. Eng. Sci.* **376**(2133), 20180087 (2018). <https://doi.org/10.1098/rsta.2018.0087>
15. Stahl, B.C.: Responsible innovation ecosystems: ethical implications of the application of the ecosystem concept to artificial intelligence. *Int. J. Inf. Manag.* **62**, 102441 (2022). <https://doi.org/10.1016/j.ijinfomgt.2021.102441>
16. Fosso Wamba, S., Bawack, R.E., Guthrie, C., Queiroz, M.M., Carillo, K.D.A.: Are we preparing for a good AI society? A bibliometric review and research agenda. *Technol. Forecast. Soc. Change* **164**, 120482 (2021). <https://doi.org/10.1016/j.techfore.2020.120482>

17. Boden, M.A.: AI: Its Nature and Future. Oxford University Press (2016)
18. Bostrom, N.: Superintelligence: Paths, Dangers. Oxford University Press, Strategies (2014)
19. Musiani, F.: Governance by algorithms. *Internet Policy Rev.* 2(3), 2013. Accessed: Feb. 03, 2023. <https://policyreview.info/articles/analysis/governance-algorithms>
20. Shneiderman, B.: Design lessons from AI's two grand goals: human emulation and useful applications. *IEEE Trans. Technol. Soc.* 1(2), 73–82 (2020). <https://doi.org/10.1109/TTS.2020.2992669>
21. Schölkopf, B., et al.: Toward causal representation learning. *Proc. IEEE* 109(5), 612–634 (2021). <https://doi.org/10.1109/JPROC.2021.3058954>
22. Kumar, A.: National AI policy/strategy of India and China: a comparative analysis. *Res. Inf. Syst. Dev. Ctries.* 2021.
23. 'Digital India: For digitally empowered society and knowledge economy| National Portal of India', India.gov.in. Accessed: Mar. 23, 2023. <https://www.india.gov.in/spotlight/digital-india-digitally-empowered-society-and-knowledge-economy#tab=tab-1>
24. UNESCO. Recommendation on the Ethics of Artificial Intelligence. 2021. Accessed: Mar. 23, 2023. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
25. Wakunuma, K., de Castro, F., Jiya, T., Inigo, E.A., Blok, V., Bryce, V.: Reconceptualising responsible research and innovation from a Global South perspective. *J. Responsible Innov.* (2021). <https://doi.org/10.1080/23299460.2021.1944736>
26. 'Framework for Responsible Research and Innovation', UK Research and Innovation. <https://www.ukri.org/about-us/epsrc/our-policies-and-standards/framework-for-responsible-innovation/>
27. von Schomberg, R.: A vision of responsible research and innovation. In: Owen, R., Bessant, J., Heintz, M. (eds.) *Responsible Innovation*, pp. 51–74. John Wiley & Sons Ltd, Chichester (2013). <https://doi.org/10.1002/9781118551424.ch3>
28. Bauer, A., Bogner, A.: Let's (not) talk about synthetic biology: framing an emerging technology in public and stakeholder dialogues. *Public Underst. Sci.* 29(5), 492–507 (2020). <https://doi.org/10.1177/0963662520907255>
29. Srinivas, K.R., Kumar, A., and Pandey, N.: Report from National Case study – India. 2018.
30. Beck, L.C., Trombetta, W.L., Share, S.: Using focus group sessions before decisions are made. *N. C. Med. J.* 47(2), 73–74 (1986)
31. Wilkinson, S.: Focus group methodology: a review. *Int. J. Soc. Res. Methodol.* 1(3), 181–203 (1998). <https://doi.org/10.1080/13645579.1998.10846874>
32. Oates, C.: The use of focus groups in social science research. In: Burton, D. (ed.) *Research Training for Social Scientists*, pp. 186–195. SAGE Publications Ltd, London (2000). <https://doi.org/10.4135/9780857028051>
33. Krueger, R.A.: *Focus Groups: A Practical Guide for Applied Research*. SAGE Publications (2014)
34. Morgan, D.L.: *Focus Groups as Qualitative Research*. SAGE Publications (1996)
35. Onwuegbuzie, A.J., Dickinson, W.B., Leech, N.L., Zoran, A.G.: A qualitative framework for collecting and analyzing data in focus group research. *Int J Qual Methods* 8(3), 1–21 (2009). <https://doi.org/10.1177/160940690900800301>
36. Mishra, L.: Focus group discussion in qualitative research. *TechnoLEARN* 6(1), 1–5 (2016)
37. Smithson, J.: Using and analysing focus groups: limitations and possibilities. *Int. J. Soc. Res. Methodol.* 3(2), 103–119 (2000)
38. Baum, S.D.: Medium-term artificial intelligence and society. *Information* 11(6), 290 (2020)
39. Chakrabarti, R., Sanyal, K.: Towards a “Responsible AI”: Can India take the lead? *South Asia Econ. J.* 21(1), 158–177 (2020). <https://doi.org/10.1177/1391561420908728>
40. Digital McKinsey: Driving impact at scale from automation and AI. 2019. Accessed: Feb. 15, 2023. <https://www.mckinsey.com/~media/McKinsey/Business%20Functions/McKinsey%20Digital/Our%20Insights/Driving%20impact%20at%20scale%20from%20automation%20and%20AI/Driving-impact-at-scale-from-automation-and-AI.ashx>
41. PwC: The Potential Impact of Artificial Intelligence on UK Employment and the Demand for Skills. 2021. Accessed: Feb. 15, 2023. https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/1023590/impact-of-ai-on-jobs.pdf

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.