CONTROL OF AI: THE COLLINGRIDGE DILEMMA

Padma Iyer, Government Law College, Mumbai

ABSTRACT

The Collingridge Dilemma is a term used for the erratic behaviour posed by Artificial Intelligence (**"AI"**). A newly launched AI tool is generally difficult to control at the start due to the unpredictability of its consequences, but when such impacts come to light, it proves to be harder to mitigate. The reason for this - as elaborated below - is the *laissez-faire* approach adopted by: first, the developers right after the launch of an AI tool or chatbot; and second, the government, in its insufficient methods of regulatory oversight. Generative AI is only run by a small number of companies, *inter alia* including Google, OpenAI and Microsoft due to their extensive resources and technological knowhow. However, the issues with AI such as bias has left even the large tech companies helpless, leading to many chat-boxes namely 'Tay' and 'Grok' being shut down not too long after their launch. In a bid to tackle the growing concerns around AI, this paper firstly explores ways to reduce - if not mitigate - the dilemma posed by AI, such as Bias or Intellectual Property Infringement. The measures elaborated here are: the Precautionary Principle, Responsible Research and Innovation, and Adaptive Governance. Secondly, it is imperative to note that strict rules related to regulatory oversight cannot be implemented immediately. As AI evolves, it is inevitable to see a parallel rise in the disputes related to its usage and development. Therefore, an effective dispute resolution system must be integrated for timely resolution of disputes. This paper delves into the advantages of arbitration for determining such disputes, as opposed to national courts.

Keywords: Artificial Intelligence, David Collingridge, Dilemma of Control, Technology, Arbitration.

INTRODUCTION

The concept of Collingridge Dilemma was conceived by David Collingridge in the 1980s in his book *'The Social Control of Technology'.* A simple explanation of this idea is that when a new technology emerges, it is difficult to regulate at the nascent stages of its deployment and use because the consequences are not completely known, including the negative repercussions of its usage on the primary stakeholders i.e., the public. However, the longer we wait to enact legislation with regard to the new technology, the possibility of its control reduces and the evidence of impact subsequently increases. With respect to AI, it has been around for a considerable amount of time and it cannot be said that the risks attached to it are completely unknown. But with the launch of ChatGPT, AI has become accessible to a larger number of people. This dilemma of control with respect to the emergence of AI cannot be overcome per say, but it can be mitigated. This can be done in ways such as adopting the Precautionary Principle, Responsible Research and Innovation (RRI), or adaptive governance

Chapter 1: Control of AI

1.01. The Precautionary Principle

The Precautionary Principle effectively employs, in essence, the opposite of a *laissez-faire* approach to AI. The Merriam-Webster dictionary defines *laissez-faire* as 'a doctrine opposing governmental interference in economic affairs beyond the minimum necessary for the maintenance of peace and property rights'[1]. This is where AI safety initially comes into play. It is well known that there is a plethora of potential harms associated with AI especially in sectors of healthcare, finance, technology etc. In this regard, it is logical to take a precautionary approach. For example, if the developer of an AI meant to make medical diagnoses can prove its accuracy, and that such an innovation would not be harmful to the public, it could receive a green light. On the flip side, if it cannot be proved that the use of AI would produce a correct diagnosis, government interference would limit the use of such AI through enactment of legislation. Alternately, Castro and McLaughlin (2019)[2] argue that instead of the precautionary approach, what needs to be adopted is the innovation principle. This is where innovation is

[1] Merriam-Webster, Inc., *Laissez-faire*, https://www.merriam-webster.com/dictionary/laissez-faire (accessed 1 Mar. 2026).
[2] Daniel Castro & Michael McLaughlin, *Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence*, https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence (2019).

encouraged, but safety nets are added by the government to act as guardrails. This approach would, however, only work in a utopian world wherein a legislation can be enacted as a solution to any problem that may arise, as if such easy-to-solve problems come with a notice of its arrival so as to give legislative bodies enough time to draft the applicable legislation. It was seen how drastically and rapidly AI's perception changed in the eyes of the world with the launch Elon Musk's 'Grock' using inappropriate language.

1.02. Responsible Research and Innovation

The concept of RRI, as per the Horizon 2020 European research framework programme, seeks to reconcile the aspect of the precautionary principle and innovation principle[3]. This was designed in an attempt to include all the stakeholders in the process of research and innovation including the citizens, policymakers, NGOs, researchers, and companies to ensure the successful and safe implementation of new technology on a global level. Bourguignon (2015)[4] believes that this form of regulation is the answer to the Collingridge dilemma, since it urges researchers to question the technology being built by them, the purpose it is intended to serve, and to make policymakers aware of the potential negative consequences it may lead to. He also believes that continuous open dialogue between the various stakeholders involved will help to gauge the correctness of decisions that are being taken. While this constant stream of interaction may enhance upstream engagement and promote transparency, there are dangers of AI that even such stakeholders might not foresee. A real-life example of the dangers of AI can be seen in the Tay controversy of March 2016. Tay was a chatbot released by Microsoft on Twitter, and what started off as innocent tweets such as 'why isn't National Puppy Day every day?', within only 16 hours after its launch, Tay quickly took a dark turn and began tweeting sexist and antisemitic tweets and had to be shut down by Microsoft. The bias projected by Tay in the form of its tweets in this case poses as a risk to the emergence of AI and faces the dilemma of control that RRI may not be able to overcome completely.

1.03. Adaptive Governance

As Shirkhanloo (2025) rightly states, companies are treating AI regulation as a 'compliance checkbox'[5] meaning that with regard to regulatory compliance, only the bare minimum is being

[3] Didier Bourguignon, *The Precautionary Principle*, https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/573876/EPRS_IDA(2015)573876_EN.pdf (2015).
[4] Ibid.
[5] Anjella Shirkhanloo, *Beyond Compliance: The Case for Adaptive AI Governance*,

done so that companies can say that they have followed the rules. Moreover, the risks associated with AI are neither being checked at regular intervals, nor when the system updates, rather it is only checked at the initial stages i.e., during its launch. This static form of regulation is not going to help control the Collingridge Dilemma. If an adaptive form of governance is adopted, regulation becomes fluid. During the initial stages of the launch of a form of AI, legislation may be enacted only regarding those risks that are apparent at that time. However, as the potential for harm unfolds further, the legislation is to be constantly overseen and amended to ensure minimal scope for harm. This might be the primary responsibility from a policymaker's perspective. However, companies bear the onus of ensuring that whenever a new model is released, the system is checked and any foreseeable dangers are dealt with in a prompt manner. Further, the system must be inspected periodically and problems like bias are fixed as and when they arise. The Bletchley Declaration[6] in November of 2023 represented by thirty-three countries committed to employ risk-based policies in their regulation, including tools for testing safety, developing scientific research, and laying out appropriate rules for regulatory oversight. Risk-based policies means governance proportional to the level of risk involved. Essentially, higher risk = stricter oversight. The European Union (EU) AI Act and the United States (US) Colorado AI Act both have adopted a risk-based approach specifically in the area of education, employment, finance, housing, healthcare and law.

Chapter 2. Resolution of Disputes Arising Out of AI

The Chief Justice of the Singapore Supreme Court, Hon'ble Chief Justice Sundaresh Menon, addressed this dilemma at his keynote address *'Reimagining Law and Technology'* at the 2025 Tech Law Fest. He rightly stated that technology is developing at an extraordinary rate, and the legal sector cannot afford to fall behind[7]. In light of this, jurisdictions across the globe must integrate efficient dispute resolution mechanisms which quickly resolves AI-related disputes such as use of copyright in training datasets and dominance of gatekeepers in generative AI. Alternative Dispute Resolution, particularly arbitration, has already begun determining technological disputes related to cryptocurrency and bitcoin. Therefore, it would be well-suited

https://iapp.org/news/a/beyond-compliance-the-case-for-adaptive-ai-governance (2025).

[6] Department for Science, Innovation & Technology, *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 (2023).

[7] Supreme Court of Singapore, *Chief Justice Sundaresh Menon: Keynote Address at the TechLaw.Fest 2025*, ¶ 2, https://www.judiciary.gov.sg/news-and-resources/news/news-details/cj-sundaresh-menon--keynote-address-at-the-techlaw.fest-2025 (2025).

to also adjudicate the unique challenges posed by AI. These advantages of arbitration, over time-consuming and costly court proceedings, are enumerated below:

2.01. Appointment of arbitrators with relevant expertise

Unlike courts, parties to an arbitral proceeding have significant amount of party autonomy to nominate an arbitrator based on the nature of the dispute. The London Court of International Arbitration (LCIA) Rules 2020, under art. 5.7, prohibit the parties from appointing arbitrators under the arbitration agreement[8]. However, under art. 5.9, due regard is given to any criteria of selection agreed by the parties[9]. Thus, parties are free to appoint arbitrators with the necessary technical expertise. It is also interesting to note that in the Singapore International Arbitration Centre (SIAC), appointments of arbitrators are made on the basis of their expertise, attributes, and track record. SIAC already has a specialist panel of arbitrators for restructuring and insolvency, as well as intellectual property disputes[10]. Therefore, arbitrators well-versed in the field of AI would be appointed. This protects parties from the judges of domestic courts who would not be able to accurately decide AI-related disputes, for want of knowledge in this field.

2.02. Cross-border AI deployment

National courts are restricted to the statutory framework of that particular territorial jurisdiction only. However, with the transnational nature of AI deployment and usage, international arbitration provides for a unified dispute resolution forum. Moreover, an arbitral award can be enforced and executed in different jurisdictions that are signatories to the New York Convention.

2.03. Confidentiality

The field of AI development is inherently sensitive. Generative AI companies such as OpenAI and Meta would not want to reveal their AI training methods and technological know-how to the public or media during the course of a dispute. For this reason, arbitral proceedings consist of confidentiality and privacy rules. Under SIAC and JAMS Rules, these rules extend even up to the existence of arbitral proceedings. These privacy standards differ vastly from courts,

[8] *LCIA Arbitration Rules* art. 5.7 (1 Oct. 2020), London Court of International Arbitration.
[9] *LCIA Arbitration Rules* art. 5.7 (1 Oct. 2020), London Court of International Arbitration.
[10] Singapore International Arbitration Centre, *SIAC Specialist Panel of Arbitrators for Intellectual Property Disputes*, https://siac.org.sg/siac-specialist-panel-of-arbitrators-for-intellectual-property-disputes (accessed 24 Apr. 2026).

wherein anyone would be allowed to spectate the proceedings, and judgements are published almost immediately on the relevant court's website.

CONCLUSION

To summarize, the Precautionary Principle ensures that before one even begins with the development of technology, it is proved that its harms will be minimal and controllable. RRI promotes open dialogue and transparency in the innovation process, and adopting adaptive governance forces companies to not take a relaxed approach to the risks posed by AI. The approaches stated above cannot be treated as a straitjacket formula to ensure that the dilemma of control by Collingridge can be easily mitigated, it is only ways to ensure that the evidence of impacts does not increase exponentially to a point where regulation becomes practically impossible. Lastly, in the event that the government and developers are unable to keep pace with the evolving nature of AI, it is imperative for a dispute resolution mechanism, particularly arbitration, to be in place which efficiently resolves the disputes in a timely manner, and is also able to navigate the web of cross-border disputes.

REFERENCES

1. Anjella Shirkhanloo, *Beyond Compliance: The Case for Adaptive AI Governance*, https://iapp.org/news/a/beyond-compliance-the-case-for-adaptive-ai-governance (2025).

2. Daniel Castro & Michael McLaughlin, *Ten Ways the Precautionary Principle Undermines Progress in Artificial Intelligence*, https://itif.org/publications/2019/02/04/ten-ways-precautionary-principle-undermines-progress-artificial-intelligence (2019).

3. Department for Science, Innovation & Technology, *The Bletchley Declaration by Countries Attending the AI Safety Summit, 1–2 November 2023*, https://www.gov.uk/government/publications/ai-safety-summit-2023-the-bletchley-declaration/the-bletchley-declaration-by-countries-attending-the-ai-safety-summit-1-2-november-2023 (2023).

4. Didier Bourguignon, *The Precautionary Principle*, https://www.europarl.europa.eu/RegData/etudes/IDAN/2015/573876/EPRS_IDA(2015)573876_EN.pdf (2015).

5. *LCIA Arbitration Rules* art. 5.7 (1 Oct. 2020), London Court of International Arbitration.

6. *LCIA Arbitration Rules* art. 5.9 (1 Oct. 2020), London Court of International Arbitration.

7. Merriam-Webster, Inc., *Laissez-faire*, https://www.merriam-webster.com/dictionary/laissez-faire (accessed 1 Mar. 2026).

8. Singapore International Arbitration Centre, *SIAC Specialist Panel of Arbitrators for Intellectual Property Disputes*, https://siac.org.sg/siac-specialist-panel-of-arbitrators-for-intellectual-property-disputes (accessed 24 Apr. 2026).

9. Supreme Court of Singapore, *Chief Justice Sundaresh Menon: Keynote Address at the TechLaw.Fest 2025*, ¶2, https://www.judiciary.gov.sg/news-and-resources/news/news-details/cj-sundaresh-menon--keynote-address-at-the-techlaw.fest-2025 (2025).