MANOHAR PARRIKAR INSTITUTE FOR DEFENCE STUDIES AND ANALYSES
मनोहर पर्रिकर रक्षा अध्ययन एवं विश्लेषण संस्थान

MP-IDSA
Issue Brief

Dealing With the Artificial General Intelligence Brahmastra

Vivek Chadha

February 20, 2026

Summary

The progression of Artificial Intelligence (AI) is being viewed with excitement and caution. This sentiment becomes even more acute when the future of AI, in the form of a self-correcting, self-improving, autonomous Artificial General Intelligence (AGI), is discussed. In many ways, AGI will emerge as the Brahmastra of modern times. Are there lessons that can be learnt from the past to help regulate the future?

Introduction

The Brahmastra is considered the most powerful weapon in the Indian epics, such as the Ramayana and the Mahabharata. A handful of warriors had mastered its employment in battle. This included Rama, Drona, Arjuna and Ashwatthama, to name a few. The knowledge of the weapon was imparted to a select warrior considered worthy of the responsibility. Drona, Ashwatthama's father, chose his son as a recipient. This, as events proved, was a wrong choice.

Ashwatthama not only employed the weapon in anger against an innocent child in his mother's womb, but he also failed to exercise control when asked to withdraw it. The one wielding the weapon was also prohibited from using it against an adversary unfamiliar with its impact and incapable of neutralising it. Yet, Acharya Drona used the weapon during the 18-day war against the Pandava army. This led Dhrishtadyumna, Droupadi's brother and the commander of the Pandava Army, to accuse him of *adharma*. After Drona was killed by deception, Dhrishtadyumna justified the act, citing Drona's violation of the rules of war.

Is Artificial General Intelligence (AGI) the new-age Brahmastra? Can lessons derived from the Mahabharata guide us in dealing with a technology that might come to represent unmatched power and influence?

Road to AGI

If the Mahabharata represents the pinnacle of philosophical and strategic thought among all epics, with its timeless wisdom, Artificial General Intelligence (AGI) will sooner rather than later become that one source that will supersede the brightest room full of Nobel Laureates. Much like the *Brahmastra*, it will represent the pinnacle of technological prowess, capable of protecting against the gravest threat and, simultaneously, causing widespread upheaval. Together, both describe the acquisition, employment and need for control over immeasurable power that can have a profound impact on their respective eras.

In its earlier avatar, AI systems have already beaten the best players at the most complex game—Go.[1] AI ensured that, in 2024, the Nobel Prize in Chemistry was awarded, among others, to a computer scientist, Demis Hassabis, for using AlphaFold to solve a five-decade-old challenge in biological computational protein design.[2] AI is already proving to be a blessing for endeavours in the medical sciences,

[1] **"World's Best Go Player Flummoxed by Google's 'Godlike' AlphaGo AI"**, *The Guardian*, 23 May 2017.

[2] **"They Cracked the Code for Proteins' Amazing Structures"**, The Nobel Prize, 9 October 2024.

possibly offering a cure for cancer in the future. Why then should its lightning-fast evolution to AGI become a concern?

The emergence of Artificial Intelligence (AI) and its aspirational eventuality, AGI, poses a dilemma not only for security-related issues but also for the future of humanity across every sphere of economic and social life. This dilemma stems from the immense power that AGI will represent. More importantly, it may no longer remain a tool in human hands and under human control. Instead, it could become an autonomous machine that replaces human decision-making on critical issues.

AI's impact is already being felt on more routine job responsibilities. Anthropic CEO Dario Amedei warned that half of all white-collar jobs in the next five years will be wiped out by AI.[3] Similarly, Google in February 2026 reportedly announced voluntary exit packages to more fully "commit to its AI-driven future", as per its Chief Business Officer.[4] Beyond AI's social and economic impact, which is becoming evident, its future manifestation as AGI emerges as the bigger concern.

At this juncture, in addition to being more 'intelligent' than humans, it will also develop the ability to self-learn. This implies that AGI will teach itself to improve rapidly, which seems a positive step, only until we realise that, having lost control over the trajectory of this learning curve, its future direction could become hazy and unpredictable.[5] This includes its ability to self-preserve, replicate and look after what it perceives as its self-interest.

In other words, AI, which until now served human interests, or at least what was perceived as human interest, will now have an 'independent' view based on its self-learning journey that would bring about alterations in its security protocols on self-preservation, possibly through replication, deceit, deception, possibly, along with the future of the planet as AGI or its successors saw in the best interests of all beings.

In an ideal scenario, this suggests that AGI will help create a better world than the parochial, self-serving interests of humans. On the contrary, even a slight possibility that humans become the 'other' or that the journey towards reaching AGI and beyond is guided by ideas of domination, control and power, and that its result could become catastrophic.[6] And here lies the dilemma of AI's fast progression towards an uncertain future. How can this uncertain future be made more predictable? And does

[3] **"Uncontained AGI Would Replace Humanity"**, *AI Frontiers*, 19 August 2025.

[4] **"Google to Employees: You can go for our Voluntary Exit Plan, if your are not enjoying..."**, *The Times of India*, 11 February 2026.

[5] Lance Eliot, **"Forewarning That There's No Reversibility Once We Reach AGI and AI Superintelligence"**, *Forbes*, 2 July 2025.

[6] Anthony Aguirre, **"Uncontained AGI Would Replace Humanity"**, *AI Frontiers*, 18 August 2025.

ancient wisdom hold any lessons in this regard? This brings us back to the Mahabharata and its lessons.

Lessons from the Mahabharata

The Kuru dynasty was guided by some of the finest minds advising the king, Dhritarashtra. This included Bhishma, Vidura, Sanjaya and Kripacharya, among several others. Yet, they witnessed a horrific war, causing widespread death and destruction. What drove the dynasty and, by implication, some of the finest minds in history that came up with the most profound intellectual guidance, to a course programmed for self-destruction?

The simple answer to this question is greed for power and possessions, accompanied by the belief that the gamble for victory is well worth the risk, irrespective of the collateral damage. The Mahabharata saw Duryodhana refusing to heed to sane advice. It also witnessed influential figures who had the power to stop the impending carnage, dither rather than intervene. Once the war began, despite agreements to the contrary, the rules of combat were sidelined in the quest for victory. The downward spiral culminated in the decision to employ the *Brahmastra*, the most potent weapon of its time, which could only be controlled by a select few. Ashwatthama fired it in frustration and anger. In retaliation, Arjuna also fired the weapon. Realising the devastating implications, Arjuna recalled it. However, Ashwatthama did not possess the same skill. The weapon destroyed its target and needed divine intervention to offset its deathly impact.

This brief account from the Mahabharata holds invaluable lessons for humanity. The ability to have a profound influence has repeatedly emerged in several forms, creating a perceptible power differential. This power, derived from economic, military, or knowledge-based advantage, had the inherent capacity to serve society's larger good while simultaneously unleashing a destructive streak. The Mahabharata suggests that *Dharma* was a guide for rulers and society. It created what can today be called a rule-based order. The wise were employed to interpret dilemmas that saw conflicts of interest or lacunae in understanding. Power was bestowed upon those who displayed the maturity and sagacity to uphold the tenets of *dharma*. Imparting the knowledge of the *Brahmastra* to Arjuna was therefore justified, and to Ashwatthama, a case of misjudgement in human behaviour.

The Challenge of AGI

The world is experiencing a period that precedes the firing of the *Brahmastra*, the equivalent of AGI. The onset of AGI is likely to be accompanied by a degree of

autonomy that may take away human control over its application. In 2023, the foremost and finest scientists, including some who were responsible for its present form and possibly its future, signed a one-line open letter warning of AI's potential consequences. The statement read, "Mitigating the risk of extinction from AI should be a global priority alongside other societal-scale risks such as pandemics and nuclear war."[7]

This is very similar to repeated caution from those who better understand AI's developmental cycle. There may be differences in predicting the date of AGI's final introduction. However, there is no difference regarding its eventual realisation. Unfortunately, unlike the Mahabharata, the naysayers do not have the control to stop the unregulated use of AGI. Nor will divine intervention come to the aid of the AGI *Brahmastra*, especially if an Ashwatthama fires it.

The current trajectory of AGI development is moving so quickly that its achievement is inevitable. This journey will reveal discoveries and capabilities once considered science fiction. It could simultaneously create challenges that could get obfuscated by the hunger for scientific discovery and the quest for strategic dominance. Under such circumstances, how can it be ensured that Ashwatthama does not get access to the *Brahmastra*?

The easier answer is to establish the bottom line for its development. AGI should not come to fruition without someone taking responsibility and accountability for its potential consequences. The more difficult challenge is making this sentiment work in the real world.

Regulating AGI

There has been a longstanding debate around regulating AI.[8] The concept of regulation can be misconstrued if it is not seen in the right context. In an ideal world, AI should be democratised, ensuring its benefits and opportunities are available to all while safeguarding society from its adverse effects. If this desirable objective is derived from the Mahabharata example quoted above, it suggests that AGI's use and control should be guided by the principle of *dharma* (righteousness) for the greater benefit of society (to help protect and improve prosperity). Its potential for abuse is minimised (building failsafe mechanisms through processes and accountability). How does this translate to the nation's vision for AI, and why is India suitably placed to be part of this endeavour?

[7] **"Statement on AI Risk"**, Centre for AI Safety.

[8] Dasharathraj K. Shetty et al., **"Analyzing AI Regulation Through Literature and Current Trends"**, *Journal of Open Innovation: Technology, Market, and Complexity*, Vol. 11, No. 1, March 2025.

Prime Minister Narendra Modi inaugurated the AI Impact Summit on 19 February 2026. He outlined India's M.A.N.A.V. vision for AI. He explained it as:

> Moral and Ethical Systems: AI must be based on ethical guidelines.
>
> Accountable Governance: Transparent rules and robust oversight.
>
> National Sovereignty: Data belongs to its rightful owner.
>
> Accessible and Inclusive: AI must not be a monopoly, but a multiplier.
>
> Valid and Legitimate: AI must be lawful and verifiable.[9]

If this vision is realised, it suggests the way ahead. This includes possible safeguards for the development of AGI. The vision reinforces righteousness as its core ethical foundation, accountability and legitimacy as the basis for creating guardrails, sovereignty, irrespective of the dispersion of capabilities across the international arena, and, finally, its development for universal good. This is easier to articulate than to achieve in an environment of acute competitiveness.

One way to achieve these objectives is to build international consensus around this sentiment and decentralise the pursuit of this vision through an international organisation on the lines of the Financial Action Task Force (FATF).[10] The FATF and its affiliate bodies are a good example of how nations can come together to address a common cause, such as terrorist financing and money laundering. It is an inter-governmental body that lays down guidelines and conducts visits to evaluate the robustness of countries' existing systems and their implementation.

AI and its progression into AGI are far more decentralised for a single body to regulate. However, an empowered task force working to control the adverse effects of AI and AGI can provide the platform to pursue considered initiatives. It can also become the repository for inputs from domain experts who have been contributing individually, cautioning against the fallout of AGI, especially if it goes rogue.[11] This can be similar to FATF-like regional bodies and affiliated organisations, such as the Egmont Group, which coordinates the Financial Intelligence Units.

India is ideally placed to take the lead for this initiative. Besides being the most populous country and therefore most likely to be affected by the outcome of the AI revolution, India is not a part of the great power rivalry. India's diversity can also serve as an ideal test bed for implementing AI's benefits. The country is widely accepted as a representative of the Global South, making its voice both

---

[9] **"PM Inaugurates India AI Impact Summit 2026"**, *PMINDIA*, 19 February 2026.

[10] **"Who We Are"**, FATF.

[11] For a more detailed perspective on AI control, see Mustafa Suleyman and Michael Bhaskar, *The Coming Wave*, The Bodley Head, London, pp. 225–280.

representative and responsible to the wider cause of humanity. At the same time, India has the intellectual heft and knowledge base to lead the effort, despite its obvious complexity.

Conclusion

The world is poised at a critical juncture. A winner-takes-all approach can only lead to a *Mahabharata* that heralded the onset of *Kali Yuga*. It is high time to build consensus on the evolution of AI and AGI. This is in the interest of all nations, regardless of their level of AI advancement. It is also time to put in place the necessary international structures. The contours of such an organisation can work to mitigate a battlefield devoid of victors. This is a war with ourselves that will only create losers unless the process is guided to reap AI's immense benefits. And benefits there are in equal measure.

About the Author

Col. Vivek Chadha (Retd.) is Senior Fellow at the Manohar Parrikar Institute for Defence Studies and Analyses, New Delhi.

Manohar Parrikar Institute for Defence Studies and Analyses is a non-partisan, autonomous body dedicated to objective research and policy relevant studies on all aspects of defence and security. Its mission is to promote national and international security through the generation and dissemination of knowledge on defence and security-related issues.

Disclaimer: Views expressed in Manohar Parrikar IDSA's publications and on its website are those of the authors and do not necessarily reflect the views of the Manohar Parrikar IDSA or the Government of India.

© Manohar Parrikar Institute for Defence Studies and Analyses (MP-IDSA) 2026

Manohar Parrikar Institute for Defence Studies and Analyses
1, Development Enclave, Rao Tula Ram Marg
New Delhi 110 010 India
T +91-11-2671 7983 F +91-11-2615 4191
www.idsa.in
Twitter @IDSAIndia
www.facebook.com/ManoharParrikarInstituteforDefenceStudiesAnalyses