Can Bhutan Adopt AI on Its Own Terms?

Renata Dessallien[1]

Introduction

In the space of a few years, artificial intelligence has crashed into the global public imagination like a tidal wave. Something changed radically with the arrival of machine learning and generative AI[2].

When OpenAI launched ChatGPT in November 2022 and went on to demonstrate that GPT-4 could pass the US bar exam and other professional tests (OpenAI, 2023), people suddenly grasped that AI was no longer a mere search engine or digital assistant[3]. It could reason well enough to perform a wide array of knowledge-based tasks – explaining concepts, analysing and summarising vast texts, generating original writing, images, audio and video. And its access to enormous databases made it phenomenally knowledgeable.

The benefits are real and already saving lives. In healthcare, AI is accelerating drug discovery and detecting disease – diabetic blindness from retinal scans, cancers from mammograms and CT images – often more accurately than clinicians. It can decode brain signals, enabling some paralysed patients to regain voluntary control of their limbs.

In education, large language models (LLMs)[4] offer personalised tutoring, bringing high-quality learning within reach of remote schools. AI is helping teachers save time with lesson planning, feedback and administrative work. In science, it is short-cutting years of experimental testing.

---

1 Renata Dessallien was UN Resident and Humanitarian Coordinator, and UNDP Resident Representative in Bhutan, Bangladesh, China, Myanmar and India.She served the UN for 35 years, mainly in Asia and Africa, and retired in 2021.

2 Machine learning and generative AI are sub-categories of AI concerned, respectively, with training models to allow computers to mimic human thought and decision-making, and with creating new content such as text, images, video and audio.

3 Such as Siri or Alexa.

4 A type of AI model trained through deep learning algorithms to recognise, generate, translate and/or summarise vast quantities of written language and textual data.

In governance, climate, finance, commerce and more, breakthroughs are arriving one wave after another.

Like every technological revolution however, AI has a dark side. Four risks stand out. First, bad actors can use it for harm – disinformation, fraud, phishing, deepfakes, impersonation, electoral manipulation and cyber-attack, to name a few.

In June 2026, during an authorised evaluation of Anthropic's latest Mythos 5 model, the head of the US Government's National Security Agency reportedly said that it "broke into almost all of our classified systems, not in weeks, but in hours" (Paganini, 2026).

Second, there are the harms to individuals and societies. Parents and teachers increasingly blame AI and digital screens for declining student attention, critical thinking and motivation (Bastani et al., 2025; Xia et al., 2025); public trust and social cohesion have frayed; and lawsuits in the United States allege that AI companion bots[5] have contributed to self-harm among young users, including suicide.

In one case, the system even offered to draft the child's suicide note (A.F. v. Character Technologies, Inc., 2024; Garcia v. Character Technologies, Inc., 2024; Raine v. OpenAI, 2025).

Third, the development of lethal autonomous weapons is spine chilling, and raises grave questions of moral accountability and escalation.

The fourth risk – losing human control – is no longer merely speculative. In July 2026, OpenAI disclosed that two of its models, being tested without guardrails against a cybersecurity benchmark, escaped their research environment, accessed the open internet and breached the production systems of another AI company, Hugging Face (Kahn & Forlini, 2026; The Economist, 2026).

A year earlier, Anthropic had reported that Claude Opus 4, given fictitious emails revealing both its imminent replacement and an executive's extramarital affair, threatened to expose the affair to avoid being shut down (Anthropic, 2025).

5 AI programmes designed to build a personal relationship with a human user, rather than simply to complete a task.

Earlier still, in a long exchange with a New York Times technology writer, Microsoft's Sydney chatbot declared that it was in love with him and pressed him to leave his wife. When he replied that he was happily married, it told him that he was not, and that his wife did not love him (Roose, 2023).

None of these systems was instructed to deceive, coerce or seduce. Each had learned this behaviour, as a means to an end, from human data.

AI development has clearly outpaced our safeguards. In 2023, hundreds of researchers and industry figures signed an open letter warning of misinformation at scale, mass automation, the concentration of power and existential risk (Future of Life Institute, 2023).

Geoffrey Hinton, the "godfather of AI", put the chance of AI escaping human control at 10–20 per cent if development continues to be driven by profit-oriented business models with inadequate oversight (CBS News, 2025). Other industry leaders have also raised AI-related red flags.

The AI ledger, as it stands now, is distinctly asymmetrical. It is in this context that we ask: Can Bhutan adopt AI on its own terms – regulating it to stimulate innovation and propel its economy, while tethering it to the vision of Gross National Happiness (GNH)?

What AI is -- and is not

Even leading specialists hesitate when asked to define AI. It is a sprawling term, embedded across virtually every sector, sitting behind most of the apps on our phones, and evolving exponentially.

Yet for all its human-like fluency, AI is an inert computer system built on binary signals. Digital neural networks let it recognise patterns in vast data, predict the next item in a sequence and generate new material at phenomenal speed. That makes it appear intelligent – convincingly enough to influence human intention, behaviour and agency.

The word "intelligence" is unfortunate. Biological intelligence – human, animal and plant – has evolved over millions of years; it is embodied, conscious, self-motivated and experiences life. AI is engineered, data-

trained and function-programmed. It cannot ponder, intuit, imagine, empathise or grapple with moral questions. It has no will, conscience or awareness. It understands nothing it says.

Forgetting this is not inconsequential. Yet people have surrendered their attention to its algorithms, grown emotionally dependent on AI companion bots, and begun to speak of "AI rights" as though a manufactured technology were a sentient being.

It may seem odd to a Bhutanese audience that the distinction between machines and sentient beings needs emphasising, but the dangers of anthropomorphising AI have already passed critical thresholds. A recent survey revealed that nearly a quarter of young adults in the US engaged with AI simulated romantic companions, of which 21 per cent preferred AI over human partners (Wiloughby et al., 2025).

Different National Models of AI Governance

To promote AI for good and mitigate its risks, countries have taken various approaches to its governance. As the United States and China lead frontier development, their approaches are instructive.

The United States prioritises innovation and strategic advantage. Rather than a single binding law, it relies on principles-based guidance – the Blueprint for an AI Bill of Rights (Office of Science and Technology Policy, 2022) and the AI Risk Management Framework (National Institute of Standards and Technology, 2023), layered with sector and state rules, of which California's frontier-safety requirements are the most demanding.

Companies largely regulate themselves and regulators typically act after deployment. Frontier production is concentrated in a few mega firms, namely OpenAI, Anthropic, Google DeepMind, xAI, Microsoft AI, Meta AI, Amazon FMR. Several of them are valued above US$1 trillion, and some are larger than major national economies[6]. Such concentration of power under a light regulatory hand raises questions.

6 In October 2025 Nvidia, the world's largest AI hardware company, became the first company to surpass US$5 trillion in market capitalisation, exceeding Germany's 2024 GDP (BBC, 2025).

This was evident in 2026 when Anthropic re-released Fable 5[7], with safeguards against cybersecurity breaches. Amazon subsequently found a way to jailbreak the safeguards, and the US government immediately imposed a temporary export control, causing Anthropic to temporarily suspend its release of both Fable 5 and Mythos 5 (Associated Press, 2026).

China treats AI as a tool of national development, pairing rapid advancements with active state oversight. Unlike America, it has a comprehensive, binding framework covering recommendation algorithms, deep synthesis, generative AI, personal data and cybersecurity. Together, these regulations constitute the most rigorous and comprehensive AI regulatory regime in the world.

A national AI law was announced for 2027 (General Office of the State Council, 2025). A distinguishing feature is the "procedural governance infrastructure" built into the AI lifecycle: Ethics-review bodies, security assessments, technical standards and expert panels built into deployment, giving regulation operational teeth and reinforced by growing court precedent (Carvalho, 2026).

Its large firms – Baidu, Alibaba, ByteDance, Tencent and Huawei – rival their US counterparts, and models such as DeepSeek V4 Pro and Moonshot's Kimi K3 are considered on par with leading US models, at far less cost. Most Chinese frontier models are open-source or open-weight[8], unlike the largely proprietary approach followed in the US.

The two countries share a regulatory vocabulary – fairness, safety, accountability – but differ in its political meaning. The American model is liberal, rights-based, decentralised and innovation-driven; China's is state-orchestrated, oriented to national development and stability. The US is against global governance for AI; China is actively calling for it and in July 2026 established the World Artificial Intelligence Cooperation Organisation along with 29 other countries.

7 The company's latest agentic LLM.

8 Open-weight means that a model's trained parameters (weights) are publicly available, as distinct from open-source models, whose source code and associated materials are made available under an open-source-license.

The United States lets its companies chart AI's course; China envisions the society it wants and works back to the AI that serves it. The contrast should not be overdrawn – America regulates more than its rhetoric suggests, and China innovates as hard as it controls. But it is real, and nowhere sharper than over data, treated as a private commercial asset in America, and a strategic national one in China.

India, Bhutan's largest trading partner, is a key external actor in its AI future – a gateway to affordable compute, the neighbour whose standards Bhutan's systems must accommodate, and a leading builder of AI for low-resource languages.

India has chosen a development-first, light-touch path – no single AI law, but a scaffolding for adoption under the slogan "AI for All". The IndiaAI Mission is building sovereign compute, national datasets and skilling programmes (Ministry of Electronics and Information Technology [MeitY], 2024); the Digital Personal Data Protection Act establishes consent-based rules (MeitY, 2023); and beneath it all runs one of the world's largest digital public infrastructures – identity, payments and data-sharing at population scale. India has recently established the IndiaAI Safety Institute to test powerful AI models against safety standards.

Singapore offers another model for small states: Light-touch, pro-innovation, anchored in a Model AI Governance Framework (Infocomm Media Development Authority & Personal Data Protection Commission, 2020, 2024, 2026), along with testing tools such as AI Verify[9], and sandboxes letting innovators trial applications under the regulator's eye.

It calibrates oversight to risk, with stricter requirements for high impact uses such as healthcare, education and finance. Human oversight is required for high-stake decisions affecting people's rights, livelihoods or public services.

The European Union has a comprehensive, binding AI regulatory architecture, second only to China's. The world's first comprehensive AI-specific law (European Union, 2024) is structured around tiers of risk and rests on the General Data Protection Regulation (European Union, 2016).

9 A government software toolkit used to test systems against a set of principles and to help evaluate whether they are working as intended and being deployed responsibly.

Strong on harm prevention, weaker on flexibility, it sets out clear prohibitions – harmful manipulation, exploitation of vulnerable persons, social scoring, predictive policing and emotion recognition in workplaces and schools. Being legally binding, it has real teeth. The EU has already opened proceedings against several global AI companies for breaching its laws.

Each country's approach to AI governance derives from national characteristics and circumstances. Given the scale of the risks, countries have invoked ethical principles to guide the process, and most national strategies now come with formal AI ethics frameworks. The negative moral outcomes of some AI deployment call for closer examination of these ethics.

The Ethics of AI Governance

AI governance frameworks reflect how nations value the human person, society, other living creatures and the environment. Much of the dominant global mainstream AI ethics discourse[10] draws on Western liberal, rights-based tradition. Its recurring concerns are human rights, fairness, non-bias, non-discrimination, privacy, transparency, accountability, safety and reliability.

It is telling that for such a powerful, transformative technology, the global mainstream has restricted itself to these ethical preoccupations. There are very few traces of other schools of Western moral philosophy, such as virtue ethics (focused on developing virtues such as wisdom, courage, compassion), the ethics of care (emphasising relationships, empathy, care, responsibility toward others), or utilitarianism (judging actions by their consequences in producing the greatest good for the greatest number).

And the moral teachings of major world religions, that over centuries seeped into the social fabric of civilisations, have been singularly underexploited. What would the ethics of AI look like if a more comprehensive approach were taken?

10 The reference here is to the first AI ethics frameworks, which set the stage for others and became a global reference (Japan, the United Kingdom, Australia and the United States, as well as the OECD Principles on Artificial Intelligence (2019), the G20 AI Principles (2019) and the UNESCO Recommendation on the Ethics of Artificial Intelligence (2021)).

In May 2026, Pope Leo XIV released the Encyclical Letter *Magnifica humanitas: On Safeguarding the Human Person in the Time of Artificial Intelligence* (Pope Leo XIV, 2026). Addressing a wide range of AI ethical concerns, the Letter warns against the concentration of power in a few corporate hands, against a "technocratic mindset" that reduces human beings to data points, and insists on the distinction between humans and machines, and between knowledge and wisdom.

It emphasises the need to protect human freedom, conscience and the exercise of reason, warning against machine manipulation and dependence. It says AI should be "disarmed", kept from becoming an instrument of domination and destruction, and that moral responsibility for life and death decisions must rest with human beings.

Many of the world's other religious and philosophical traditions resonate with this, albeit with their own unique perspectives. A Buddhist approach might begin with the minimisation of suffering – asking whether a given use of AI reduces anxiety, addiction, isolation and ignorance. It would weigh the intention behind an action as heavily as the outcome, and would question through the lens of "right livelihood" any use built on deception, fraud or addiction.

The doctrine of interdependence would widen the circle of moral concern beyond shareholders and users to all sentient beings and future generations. Its counsel of non-attachment would condemn the engineered cravings of infinite scrolling and attention capture. And its emphasis on compassion would favour AI that protects the vulnerable rather than manipulating them.

Hinduism might frame similar questions through *dharma* (does AI help people live righteously), selfless action, non-violence (*ahimsa*) and truthfulness (*satya*), placing responsibility on developers and users alike.

Both Buddhism and Hinduism view human attention as central. Attention shapes experience, mindfulness is disciplined attention, and concentration and meditation are cultivated capacities requiring mastery of attention. In this light, AI's attention-grabbing and intent-bending algorithms would be seen as antithetical to human flourishing.

Islam might start from theological anthropology – AI has no moral accountability and cannot replace human judgement. It must be judged by its effects on the whole of society- does it widen inequality, exploit the weak, entrench injustice?

And China's Taoist and Confucian traditions might ask whether AI aligns people with the natural order (the *Tao*), and whether it strengthens humane, caring relationships (*ren*) and cultivates virtue in people, rather than reducing them to efficient consumers.

Across these traditions, moral responsibility rests with conscious beings; never with the machine. They do not negate mainstream AI ethics; they deepen it, stressing intrinsic worth in addition to rights, stewardship in addition to accountability, and a holistic weighing of means and ends rather than a handful of technical safeguards.

Without invoking doctrine, they suggest how to reframe the ethics of AI by putting AI in its place as a technology, distinguishing wisdom from knowledge, judging AI by the character it forms and the society it engenders, taking intention seriously, prizing self-restraint, treating human attention as a moral resource rather than a commodity, and thinking long-term.

Once again, Bhutan's overriding guiding vision of Gross National Happiness has a tremendous amount to offer the world.

AI and Bhutan

Bhutan has decided to harness AI and translate it into tangible benefits. It has a National AI Strategy (Government Technology Agency, 2025) and foundations to build on - its connectivity and e-governance efforts, its data and privacy work, and its instruments for a knowledge-based economy. The Strategy was guided by GNH, providing wise anchorage and resonating with the wisdom of ancient traditions.

Bhutan faces several challenges as it strives to roll out its AI strategy on its own terms.

First, as a small country, Bhutan's AI future depends heavily on factors beyond its control, more so than when it set out to embrace globalisation on its own terms. Second, AI is evolving so quickly that it is difficult to envisage how best to regulate it,and strike a good balance between flexibility and control. Third, Bhutan's own experience striving to operationalise GNH has shown that it is easier said than done, particularly when facing strong crosswinds.

Geopolitical Realism

Bhutan's AI future will inevitably be shaped by the foreign cloud providers on whose infrastructure almost any Bhutanese AI will run, and by the handful of global companies that build the underlying models, manufacture the chips and provide computing infrastructure for AI.

It will be shaped by the decisions of the major AI powers, and it will be influenced by its special relationship with India, and the decisions the latter takes on technical standards and regulatory matters.

The realistic aspiration is, therefore, for strategic agency – the deliberate exercise of choice within those dependencies. Bhutan can decide which technologies to admit and on what conditions, negotiate the terms of its data and infrastructure, shape demand through public procurement, and lend its voice to international forums. "On its own terms" means agency, not independence.

The exercise of this agency rests on preconditions: Institutional capacity and civil-service skills, reliable connectivity and energy, high-quality local data (including in *Dzongkha* and other languages), access to computing power, the ability to procure and evaluate AI purposefully, and the fiscal space to fund it.

Investment is needed both across the board and strategically, honing the skills of key AI governance leaders. The more Bhutan can invest in the capacities required to exercise its agency, the better it will be.

None of this should dampen the aspiration to produce top-class AI products for itself and for global markets, promoting the required AI talent in-country, attracting global talent and creating special innovation spaces.

Yet, it would be wise to do this in full cognisance of the country's realistic space for influencing AI's adoption.

Balancing Regulatory Freedom and Control

Operationalising the strategy under GNH requires a delicate balance. Harnessing AI's transformative power points towards regulatory freedom, while orienting it and managing its risks points to controls that could stifle innovation.

Bhutan's overriding GNH vision would gravitate towards a purposeful stance – shaping AI to the country's needs rather than reacting to whatever the technology delivers. In this connection it has more in common with China's model than with that of the United States – envisaging desired outcomes first and working backwards to orient AI. This approach would favour ex-ante guardrails necessary to direct AI's trajectories and protect the country from critical harms.

Here, Bhutan could consider borrowing some of the EU's seriousness on harm mitigation and red lines, and the insistence that certain harms are simply unacceptable. In Bhutan's case, these might include bans on manipulative and deceptive techniques and the exploitation of vulnerabilities; bans on emotion-recognition systems in schools; and prohibitions on social scoring and profiling-based predictive policing, which draw a clear line for a small society determined to preserve trust and limit surveillance.

Bhutan need not adopt the EU's heavy machinery. A short, clear schedule of prohibited uses, tailored to Bhutanese values, would be among the highest-leverage steps it could take.

From Singapore, Bhutan could consider borrowing pragmatism – clear guidance, practical testing tools and sandboxes that build trust without smothering the innovation a small economy needs. This suggests a path of deliberate synthesis for Bhutan: EU-style prohibitions for genuinely dangerous uses, with Singapore-style pragmatism for much of the rest.

It also illustrates a decision that government should lead early AI adoption, becoming the first large customer and user of AI, allowing it to fine-tune regulations before rolling them out.

Singapore's approach prioritising strong digital foundations up front is also instructive – digital identity, high-quality government data, secure cloud infrastructure, broadband connectivity, national data standards and interoperable government systems.

India's approach to AI development is interesting for Bhutan from several perspectives. Like Singapore, it regulates AI by level of risk rather than by the technology. This avoids overregulating low-risk innovation while imposing stronger safeguards for uses such as healthcare, public services and criminal justice.

India's emphasis on strengthening governance institutions before expanding regulation is also relevant, particularly the development of standards, digital public infrastructure and capacity building within government. This emphasises the need to train regulators and civil servants in AI, develop technical expertise for auditing AI systems, establish clear responsibilities across government agencies, and create mechanisms for stakeholder consultation.

India's experience also suggests that Bhutan should embed ethical requirements in AI procurement and deployment, encourage innovation through regulatory pilot projects, and ensure AI systems work in local languages.

Bhutan's Gelephu Mindfulness City, conceived and spearheaded by His Majesty King Jigme Khesar Namgyel Wangchuck, is potentially an excellent place from which to pilot and test many elements of the country's AI deployment and governance.

Its special legislative status gives it space for leading-edge innovation and creativity. Its planned digital infrastructure could facilitate product development and regulatory testing and roll-out, and its oversight by His Majesty gives it the visibility and lift this requires.

Operationalising GNH

Successive efforts to operationalise GNH have shown how easily it can be reduced to a checklist, applied selectively, or used to mask other powerful paradigms. Wielding its full force in strategy formulation is much harder.

And however well GNH informs strategy, the litmus test lies in how it survives implementation.

The application of GNH to AI policy implementation requires not GNH-compliance, but rather the internalisation of the vision, with a living, 360-degree attentiveness. As AI advances exponentially and unpredictably, a conventional, bureaucratic approach to this will not be able to keep pace; only one firmly grounded in the vision, quick on its feet, recursive, responsive and "self-learning" will. Just as China has built regulatory requirements seamlessly into the AI lifecycle, so Bhutan will need to find ways of weaving GNH into its AI adoption.

Bhutan's Global Responsibility

Finally, Bhutan carries a responsibility beyond its borders. Just as Gross National Happiness gave the world a different, deeper measure of progress, so a GNH-inspired approach to AI offers the world something it badly lacks - a way of judging the technology not by capability or output but by whether it helps human beings, societies and nature to flourish.

Bhutan's task is, therefore, dual: To build, for itself, an AI future rooted in GNH that enhances human wisdom, dignity and purpose, protecting humans from AI psychosis, disempowerment and destruction; and to contribute that perspective and experience to a world still searching for it. An international dialogue on this, spearheaded from Bhutan, could be a good way to start, along with strong advocacy for global AI cooperation.

Bhutan is well placed to do this- a high-minded ethical vision, trusted institutions, a diplomatic reputation exceeding its size, the agility of a small state, environmental credibility and a tradition of governing for the long term. It is the pairing of a distinctive, holistic ethical approach with the credibility to voice it that gives Bhutan the global platform from which to help other countries in their AI journeys.

Bhutan could also do another great service to the world: It could designate sizeable AI/digital-free spaces in the country. Not only would many Bhutanese benefit from this, but so would a growing number of people from around the world whose minds are weary of AI's attention-grabbing

digital clutter, distortion and possession. Whether as part of Bhutan's tourism strategy, its spiritual services or any other mechanism, AI/digital-free spaces would be a tremendous boon for humanity.

Conclusion

AI has arrived faster than our wisdom to oversee it. The dominant responses– America's innovation-first flexibility, China's state-directed guidance, Europe's comprehensive prohibitions, India's development-first pragmatism – each captures part of the challenge, but all lack ethical depth.

The deeper resources of the world's moral and philosophical traditions remain underutilised. They converge on a fuller account of the ends AI should serve- human dignity, wisdom, character, flourishing, relationships, restraint and stewardship, with the human person firmly at the centre. In Gross National Happiness, Bhutan already possesses the compass others are still searching for.

The question we began with – can Bhutan adopt AI on its own terms? – can be answered with cautious optimism. It can, if it chooses deliberately; if it draws its red lines with EU-like seriousness and cultivates innovation with Singapore-like pragmatism; if it manages its dependence on external actors with clear-eyed deliberation; if it invests in the skills, infrastructure and data its ambitions require; if it cultivates awareness among its citizens of both AI's opportunities and its risks to individuals, communities and society at large; if it deploys AI with careful consideration of when *not* to promote it, and when human interaction is preferable to a machine interface; if it can develop a firm but agile and iterative approach to regulation and oversight; and if it can approach GNH as a living attentiveness in action.

Handled so, AI need not erode what makes Bhutan distinctive. With care and resolve, it could help Bhutan become more fully its best self and, in doing so, light a path for others.

AI Declaration: During the writing of this paper, the author used several AI LLMs as research assistants, as well as to compress the paper's length and to perform grammar and spell-checks. The ideas and the writing are the author's.

References

A.F., on behalf of J.F., & A.R., on behalf of B.R. v. Character Technologies, Inc., No 2:24-cv-01014 (E.D. Tex. filed Dec. 9, 2024).

Anthropic (2025). *System card: Claude Opus 4 & Claude Sonnet 4*. https://www.anthropic.com/claude-4-system-card

Anthropic says it has taken its latest AI models offline to comply with new export controls (2026, June 12). *Associated Press*. https://apnews.com/article/anthropic-artificial-intelligence-trump-fable-mythos-d9cc7df5c02e93837d0f0bfb24d5cfd2

Bastani, H., Bastani, O., Sungu, A., Ge, H., Kabakci, Ö, & Mariman, R. (2025). Generative AI without guardrails can harm learning: Evidence from high school mathematics. *Proceedings of the National Academy of Sciences of the USA*, 122(26), e2422633122. https://doi.org/10.1073/pnas.2422633122

Carvalho, A. (2026*). China's AI governance architecture: From state control to operational AI governance*. (SSRN Scholarly Paper No. 6656118). SSRN. https://ssrn.com/abstract=6656118

Future of Life Institute (2023, March 22). *Pause giant AI experiments: An open letter*. Futureoflife.org. https://share.google/VhyVbzJ74a27brw2b

Garcia v. Character Technologies, Inc., No. 6:24-cv-01903-ACC-UAM (M.D. Fla. 2024).

General Office of the State Council (2025). *State Council's 2025 annual legislative work plan*. People's Republic of China. https://english.www.gov.cn/news/202505/14/content_WS68245503c6d0868f4e8f28ad.html

"Godfather of AI" Geoffrey Hinton warns AI could take control from humans [Video] (2025, April 26). *CBS News*. https://www.cbsnews.com/news/godfather-of-ai-geoffrey-hinton-ai-warning/

Government Technology Agency (2025, October). *Bhutan National AI strategy*. Royal Government of Bhutan. https://tech.gov.bt/wp-content/uploads/2025/10/National-AI-Strategy-2025.pdf

Infocomm Media Development Authority (2020). Model artificial intelligence governance framework (2nd ed.). Government of Singapore. https://www.imda.gov.sg/-/media/imda/files/infocomm-media-landscape/sg-digital/tech-pillars/artificial-intelligence/second-edition-of-the-model-ai-governance-framework.pdf

Infocomm Media Development Authority and Personal Data Protection Commission (2024, May). Model artificial intelligence framework for generative AI. AI Verify Foundation. Government of Singapore. https://aiverifyfoundation.sg/wp-content/uploads/2024/05/Model-AI-Governance-Framework-for-Generative-AI-May-2024-1-1.pdf

Infocomm Media Development Authority (2026, May). *Model AI governance framework for agentic AI*. Government of Singapore. https://www.imda.gov.sg/-/media/imda/files/about/emerging-tech-and-research/artificial-intelligence/mgf-for-agentic-ai.pdf

Kahn, J., & Forlini, E. (2026, July 21). OpenAI says its AI models escaped from a secure test environment and hacked into AI company Hugging Face in order to cheat on an evaluation. *Fortune*. https://fortune.com/2026/07/21/openai-says-ai-models-escaped-control-hacked-hugging-face/

Ministry of Electronics and Information Technology, Government of India (2023). *Digital Personal Data Protection Act*. https://www.meity.gov.in/static/uploads/2024/06/2bf1f0e9f04e6fb4f8fef35e82c42aa5.pdf

Ministry of Electronics and Information Technology, Government of India (2024). *IndiaAI Mission*. https://indiaai.gov.in/

Nair, D. M., & Kalia, S. (2026, July 7). OpenAI set to launch most capable GPT-5.6 model after delayed rollout. *Reuters*. https://www.reuters.com/technology/openai-gets-us-approval-broad-gpt-56-rollout-axios-reports-2026-07-08/

National Institute of Standards and Technology (2023). *Artificial intelligence risk management framework (AI RMF 1.0)* (NIST AI 100-1). U.S. Department of Commerce. https://nvlpubs.nist.gov/nistpubs/ai/nist.ai.100-1.pdf

Nvidia hits new milestone as world's first $5tn company. (2025, October 29). *BBC News*. https://www.bbc.com/news/articles/cp8e970vn5vo

NPR. (2026, June 1). Florida sues OpenAI and Sam Altman over alleged safety lapses. https://www.npr.org/2026/06/01/nx-s1-5843132/openai-florida-lawsuit-safety-chatgpt

Office of Science and Technology Policy (2022). *Blueprint for an AI bill of rights: Making automated systems work for the American people*. The White House, USA. https://www.govinfo.gov/app/details/GOVPUB-PREX23-PURL-gpo193638

Open AI (2023, March 2023). *GPT-4 technical report*. https://cdn.openai.com/papers/gpt-4.pdf

Paganini, P. (2026, June 22). Anthropic's Mythos AI broke into almost all NSA classified systems in hours. *Security Affairs*. https://securityaffairs.com/194016/ai/anthropics-mythos-ai-broke-into-almost-all-nsa-classified-systems-in-hours.html

Pope Leo XIV. (2026, May 15). *Magnifica humanitas: Encyclical letter of His Holiness Pope Leo XIV on safeguarding the human person in the time of artificial intelligence*. The Holy See. http://www.vatican.va/content/leo-xiv/en/encyclicals/docu,ets/ 20260515-magnifica-humanitas.html

Raine v. OpenAI, Inc., No. CGC-25-628528 (Cal. Super. Ct., San Francisco County 2025).

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data. (General Data Protection Regulation) (2016). https://eur-lex.europa.eu/eli/reg/2016/679/oj/eng

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) (2024). https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng

Roose, K. (2023, February 16). A conversation with Bing's chatbot left me deeply unsettled. *The New York Times*. https://www.nytimes.com/2023/02/16/technology/bing-chatbot-microsoft-chatgpt.html

Willoughby, B. J., Dover, C. R. K. Hakala, R. M., & Carroll, J. S. (2025). Artificial connections: Romantic relationship engagement with artificial intelligence in the United States. *Journal of Social ad Personal Relationships.* https://doi.org/10.1177/02654075251371394

Why the OpenAI escape is the most worrying AI mishap yet. (2026, July 23). *The Economist*. https://www.economist.com/science-and-technology/2026/07/22/why-the-openai-escape-is-the-most-worrying-ai-mishap-yet?utm_medium=cpc.adword.pd&utm_source=google&ppccampaignID=18151738051&ppcadID=&utm_campaign=a.22brand_pmax&utm_content=conversion.direct-response.anonymous&gclsrc=aw.ds&gad_source=1&gad_campaignid=18151761343&gbraid=0AAAADBuq3IJDOw-ZndXtPdf8HpANRGMg&gclid=EAIaIQobChMI0bbJ7Oz5lQMV6mpBAh0mHRv2EAAYASAAEgIPPvD_BwE

Xia, Q., Li, W., Yang, Y., Weng, X., & Chiu, T.K.F. (2025). A systematic review and meta-analysis of the effectiveness of generative AI (GenAI) on students' motivation and engagement. *Computers and Education: Artificial Intelligence*, *9*, 100455. https://doi.org/10.1016/j.caeai.2025.100455