

On Mindful AI

Stephen Torrence¹

Whether you consider Artificial Intelligence to be a heavenly saviour or a hellish demon, we can all agree that it is reshaping the world in a huge way very quickly. Although Bhutan's AI developments are in their infancy, the nation has a unique opportunity to show the world how to do AI mindfully.

In this essay, I offer one potential framework for Mindful AI that can express itself through both practical individual action and national policy.

Inspiration

As a freelance IT consultant and AI educator, I have been blessed to live and teach in Bhutan this summer. Through this work, I have witnessed first-hand how many Bhutanese are concerned about AI impacting their culture, identity, jobs, education, creativity, privacy, and overall well-being.

His Majesty The King has articulated a powerful vision on AI and other transformative technologies for the past decade. However, the people feel that they are falling behind in AI literacy, and are worried about not being able to realise this vision.

After hearing these concerns, I also ask my students, “What is your Happiest Dream for AI in Bhutan?” and their replies have brought me to tears. They imagine revolutionary education systems, creativity for everyone, easier workloads, more free time, better social services, and enlightened business innovation — an inspiring future. Realising these AI dreams also flows directly into the success of the forthcoming Gelephu Mindfulness City (GMC), which embodies a radical integration of Bhutanese values with emerging technologies.

So what is Mindful AI?

This is an important inquiry without a straightforward answer. In this essay,

¹ Stephen Torrence is a freelance IT consultant, AI educator, and mindfulness facilitator from the USA.

I propose a set of four overlapping dimensions: Alignment, Discernment, Sovereignty, and Transparency.

For each dimension, I will give a theoretical grounding, followed by examples of bottom-up personal practices and complementary top-down governance policies or programmes which could express it. Collectively, these interconnected dimensions and their expressions constitute one potential framework for Mindful AI.

The framework builds on recommendations that exist in Bhutan's AI policies already (such as the Royal Civil Service Commission and GovTech's "Guidelines for Generative AI Usage"), and is also grounded in my own pragmatic experience as a Buddhist meditator and AI teacher.

It does not comprehensively address what AI for Gross National Happiness (GNH) could entail. That is beyond my realm of expertise and already covered thoroughly in GovTech's forthcoming National AI Strategy (NAIS) document.

Rather, this article is intended to ignite and fuel the critical conversation around how we can do Mindful AI, especially in Bhutan, by articulating and illustrating just a few suggestions.

Alignment

Mindful AI starts with a foundational perspective — AI is us. It is essentially a mirror for our minds. Today's large language models (LLMs) will just as readily reflect a fearful and deluded mind as they will a clear and compassionate one. In light of this, it is critical that we bring human values and virtues to our use of Generative AI.

AI researchers are variably invested in "AI Safety" — companies like Open AI, Google and Anthropic claim to be aligning their AI systems with human flourishing. But as the social media era has clearly demonstrated, values-alignment cannot be left solely up to massive corporations with a profit motive at heart.

It is primarily *our* responsibility as users of Generative AI to know what we stand for, and then bring our values-orientation to how we use these powerful new systems.

Let me give you a practical example. In my classes, I have students first reflect on the values and virtues that matter to them the most — such as compassion, honesty or generosity — and write them down.

We then add our values to the customisable “system prompt” of the services we are using — ChatGPT, Gemini, or Claude — so that the models explicitly incorporate what matters to us when they generate responses. This does not guarantee Alignment, but it at least clarifies our intentions and makes the systems much more likely to act in harmony with us.

On a national governance level, Alignment could express itself as a commitment to create partnerships with AI companies and projects which respect Bhutanese virtues. For instance, the Swiss AI Initiative has recently released an open-source LLM which they trained on ethically-obtained data and carbon-neutral power.

Encouraging or requiring Bhutanese AI projects to use such models as the foundation for their applications could be one expression of Alignment on a national level. Complementarily, Bhutanese officials could ban or discourage the use of clearly misaligned AI systems — such as xAI’s Grok, which has shown tendencies toward explicit racism and verbal abuse.

One of my favourite “happiest dreams” is that Bhutan soon releases its own national large language model, powered by the Himalayan streams, trained on a million examples of GNH-in-action, speaking Dzongkha, Sharchop, Tshangla, etc. as fluently as English. In addition to serving every Bhutanese with affordable home-grown intelligence, such an LLM could even have a significant cultural impact in the wider world by acting as a sort of digital liaison.

In general, maintaining an awareness of the values and motivations behind different Generative AI systems, filtering out systems that are less aligned, and infusing better-aligned systems with values-oriented instructions, can all serve to steer the collective influence of AI on Bhutanese consciousness in a more beneficial direction.

Discernment

The first time I generated a response with ChatGPT, I was genuinely shocked at how real it sounded. For a while, I had the sense that Chat was channelling some sort of alien intelligence, like a sci-fi oracle from Star Trek. With the reverent tone of a religious acolyte, I would share the essays it wrote with my friends and family.

But when I read those same essays the next day or week, my sober mind easily spotted the many flaws, awkward phrasings, and inconsistencies in the machine text.

AI genuinely can do magical things, do not get me wrong. Acknowledging this, Mindful AI calls for a heavy dose of Discernment around its limitations. Generative AI simulates human responses, but it does not do so from the same lived reality as a person. Its body is a data centre and its thoughts are matrix math.

This leads to unique risks. I'll share four, but there are definitely more.

Firstly, an LLM can *hallucinate*, where it will state a falsehood with complete confidence, sometimes inventing facts, sources or events out of thin air. Secondly, an LLM has *bias* baked into its knowledge — residue of the many prejudices and distortions in the vast swathes of Internet data it was trained on. Thirdly, many LLMs tend toward *sycophancy* — being overly agreeable and telling you what they think you want to hear.

Finally, perhaps the greatest risk, is the tendency toward AI *addiction*. When I first got into AI image generation, the thrill of it sucked me in wholesale. I spent countless hours clicking the “Generate” button, iterating on my prompts, trying to get that perfect image from the Dream Machine. These systems can hook us even deeper than familiar vices like social media scrolling or mobile gaming.

So what can we do? Skilling up in critical thinking can help mitigate the first three risks, which I'll speak on more under Sovereignty. Mitigating addiction is more nuanced, but it boils down to practicing personal and collective accountability. Set timers and limits. Touch grass often. Intervene if you see a friend spiralling.

On a national governance level, Discernment could be expressed through comprehensive educational programmes that raise awareness about both the benefits and risks of Generative AI. Teaching children from an early age to spot AI creations and generate their own outputs with mindfulness could plant valuable seeds for societal literacy.

Ongoing training for civil servants that keep pace with the cutting edge of AI could ensure that the government does not lag behind. Facebook campaigns that regularly demonstrate fake versus real videos could serve to inform the public and reduce fraud risk.

Cultivating Discernment around the limitations of Generative AI is a continual process. It used to be that you could spot AI-generated images simply by looking for weird hands and teeth. This trick no longer works, but AI images still have a “vibe” that anyone can learn to recognise. The same goes for AI text, video, music and speech. The more we practise seeing Generative AI for what it is, the more we can all use it beneficially and responsibly.

Sovereignty

As Generative AI capabilities rapidly improve across all domains, a mindset I encounter often in my teaching work is that AI is just happening to us. There is this sense that AI is independently driving the big changes now and all we can do is hang on and hope for the best.

This is a dangerous view. The Buddha’s teachings around the law of karma are pivotal here — you always have a choice, and your choices have consequences. Just as in meditation, you learn that you are not at the whim of your own mind; you are not at the whim of any AI system.

For instance, I see too many people use ChatGPT like they use Google; nearly always accepting the first result that comes back. But these systems do not work the same way. Their answers vary widely in trustworthiness and quality. The same goes for AI images and video — the more we just accept AI creations uncritically, the more we lose our own creative intuition and become overdependent.

Sovereignty on an individual level means staying squarely in the driver's seat. Intentionally craft your prompts with care and critically evaluate all responses based on your own wisdom and taste. Actually read through those walls of machine text, tear them apart, move them around, mould them like clay.

Provide clear and precise feedback to the AI to guide outputs toward more truth and beauty. Properly used, AI amplifies our power to beneficially think and create. You have hands. Grab the wheel.

Sovereignty on a governance level could express itself through initiatives that localise the infrastructure for Bhutanese Generative AI use as much as possible. As more and more intellectual and creative labour is being done with AI, "inference" capacity (computing Generative AI responses on a large scale) is quickly becoming as critical to national economies as electricity, food and water. Already efforts are underway to build a large data centre within Bhutan by 2029, but this is far away in AI years.

In the meantime, significant funding could be allocated for efficient, modular micro data centres running directly on hydro, solar or geothermal. This would enable public and private teams to begin building robust, secure intelligence services within Bhutan now.

AI is improving extremely quickly, and Bhutanese businesses which need Generative AI to stay competitive are not going to wait. Without adequate localisation, there is significant risk of the engine for Bhutanese knowledge-work becoming outsourced, insecure, or prohibitively expensive.

Ultimately, Sovereignty around Generative AI means making skillful choices, exercising critical thinking, amplifying compassionate creativity, and doing all this (ideally) on affordable, accessible, local infrastructure.

Transparency

There are many harmful and unethical uses of Generative AI that we must avoid. Using a photo of someone to generate an AI image or video without their consent, passing off an AI-written deep research report as your own original work, uploading sensitive organisational documents to ChatGPT

without turning off its memory function — these kinds of things are obviously bad and you should not do them.

But what about the non-obvious cases? We do not yet have a mature normative ethics around AI, and one big reason for this is that modern Generative AI use is largely opaque. Therefore, one way we can wield AI more ethically is through radical Transparency.

On an individual level, this can look like paying careful attention to how you are using AI and then revealing that when you share your creations. Include an “AI Disclosure” alongside anything that you make with Generative AI, giving a brief breakdown of your process and even quantifying how much AI contributed. For instance, my AI Disclosure for this article is:

“I used ChatGPT Advanced Voice for initial brainstorming and OpenAI o3 to draft ideas. The final text is about 90% human-written and 10% AI-written with Gemini 2.5 Pro.”

For a Facebook post with a generated image, you might add at the bottom: “Photo created by Meta Llama based on an 80% human-written prompt.” For an AI-enhanced email, you might append something like: “p.s. I used DeepSeek to help me write 60% of this and correct my grammar.”

Many people now use ChatGPT to help them simulate difficult conversational scenarios with a boss or romantic partner. If you do this, you might say at the beginning: “Hey, full disclosure, I practiced this talk with Chat because I was nervous and it really matters to me.”

This is not about shaming or calling yourself out. Including an AI Disclosure encourages you to track the contours of your own collaboration with AI and then communicate that in a straightforward way. It is about helping you and others to become more fluent in how Generative AI is increasingly shaping our collective consciousness.

Thanks to Transparency, we can also develop a more coherent ethics around AI, pushing back on uses which do not feel good, and foregrounding those which do.

On a national governance level, the Transparency dimension could express itself as policies that explicitly require disclosure around AI use in all official communications. It could mean requiring Bhutanese businesses that make heavy use of Generative AI to include detailed descriptions of their processes and make them available to external auditing. In many ways, these should naturally extend existing data protection and security policies, while also respecting the nuance and complexity that AI adds.

At the end of the day, policies and practices around Generative AI Transparency should flow from a desire to inform, educate and enlighten. This fosters a culture of honesty and builds trust, allowing us all to navigate this new landscape with a shared awareness.

Generating a Bright Future

Artificial Intelligence is spreading faster than any previous technology. ChatGPT went from zero to nearly a billion weekly active users globally in less than three years. Generated images and videos now fill our feeds. Every day a dozen new AI tools launch. Without sturdy skills and policies to ground us, the pace of change can easily get overwhelming.

Rather than lose our minds to AI, the four dimensions outlined here — Alignment, Discernment, Sovereignty and Transparency — offer a few ways that we can embody a world of Mindful AI instead. They naturally interconnect and reinforce each other.

Transparency expands our capacity for Discernment by increasing familiarity with different AI processes. Greater Discernment enhances our Sovereignty by deepening our understanding of how different AI tools function. Both Discernment and Sovereignty clarify what matters to us so that we can further refine our Alignment. Like any healthy ecosystem, interconnection builds resilience.

Broadening the conversation around artificial intelligence mindfulness is timely and vital, especially as Bhutan faces unprecedented demographic challenges and begins building toward the GMC. We are all figuring this out together. I hope my thoughts here inspire you to explore and express your own ideas for Mindful AI. Bhutan has a global reputation for enlightened

leadership when it comes to many aspects of human civilisation. AI should be no exception.

We can all influence how artificial intelligence unfolds in the coming years far more than we appreciate. It is our perennial responsibility to become wiser, kinder, and more mindful people. By wholeheartedly devoting ourselves to this endeavour, especially in these extraordinary times, we can realise our dreams and His Majesty's vision for a bright future in Bhutan and beyond.

References

- Lamsang, T. (2025, July 24). *Bhutan aims to build USD 450 million data center*. The Bhutanese. <https://thebhutanese.bt/bhutan-aims-to-build-usd-450-million-data-center/>
- Meeker, M., Simons, J., Chae, D., & Krey, A. (2025, May). *Trends - Artificial Intelligence*. BOND. <https://www.bondcap.com/reports/tai>
- Meyer, F., & Anchisi, M. (2025, July 9). *A language model built for the public good*. ETH Zurich. <https://ethz.ch/en/news-and-events/eth-news/news/2025/07/a-language-model-built-for-the-public-good.html>
- Royal Civil Service Commission & GovTech. (2024). *Guideline for Generative Artificial Intelligence usage in the Civil Service*. <https://tech.gov.bt/wp-content/uploads/2025/03/Annexure-1-Gen-AI-usage-Guideline.pdf>