ChatGPT

Executive Summary

This white paper, produced by India's Principal Scientific Adviser's office, examines the rise of **foundation models** in AI and India's strategy around them. In a nutshell, *foundation models* (large, pre-trained AI models) are transformative because the same model can be adapted to many tasks instead of training separate systems for each. This gives huge power but also concentrates influence – design choices in these base models affect many downstream applications and sectors. The paper emphasizes India's **approach**: build homegrown foundation models aligned with Indian languages, data and values, backed by public compute and datasets, while also establishing governance (guidelines, laws, safety) around their use.

Key points include:
- **Innovation Pillar:** Under the IndiaAI Mission, India launched calls for proposals and selected teams (industry+academia) to develop indigenous large language/multimodal models. By early 2026, *twelve* consortia (from startups like Sarvam AI, Soket AI, Gnani AI, etc., to IIT Bombay) have been funded to build these models on Indian data [1]. The idea is to ensure India has sovereign, open-source models for its own needs.
- **Compute Infrastructure:** A national AI Compute Portal provides massive GPU resources (tens of thousands of GPUs) at subsidized rates [2] [3]. Originally aiming for 10,000 GPUs, India has onboarded **38,000 GPUs** for public use [4]. (Extra: Globally, such infrastructure investment is huge – for context, worldwide AI spending is projected at **$2.52 trillion** in 2026 [5].)
- **Data Platform (AI Kosh):** The AI Kosh platform aggregates curated datasets and models. It hosts thousands of datasets and hundreds of models (for example, **300+ datasets** by early 2025 [6], now rapidly expanding) spanning many sectors. The goal is high-quality, India-specific data (Indic languages, local domains) so foundation models learn relevant patterns.
- **Governance Framework:** IndiaAI launched **National AI Governance Guidelines (2025)**, built around seven core "Sutras" (principles) like Trust, Fairness, Safety, etc. The approach is explicitly *risk-based*, meaning stronger rules for high-risk AI applications [7] [8]. Alongside this, the new **Digital Personal Data Protection Act, 2023** provides a legal framework for privacy and data use in AI. There are also CERT-In advisories and a new AI Safety Institute forming to handle threats and ethics.
- **AI for National Priorities:** The paper highlights focusing AI on India's challenges – e.g. healthcare, agriculture, education, local language technology, etc. (Extra: For instance, India has launched "Bharat-VISTA," a multilingual AI tool to advise farmers using local data [9]). The idea is that homegrown foundation models and applications help solve on-the-ground problems in an Indian context.

Throughout, the message is that India is building a **balanced AI strategy**: invest in advanced tech (foundational models, compute, data) while ensuring ethical guidelines, privacy laws, and alignment with India's socio-economic diversity. External data shows this aligns with global trends – India now ranks **#3** in global AI competitiveness behind only US and China [10] – and the world is pouring unprecedented money into AI (global AI spend ~\$2.5T in 2026 [5]). In summary, the document outlines how India plans to catch up with big tech in generative AI by growing its own models and ecosystem, while also keeping them safe and locally relevant.

Foundation Models: What and Why

What are foundation models? These are very large machine-learning models (often with hundreds of millions to billions of parameters) pre-trained on broad data, which can then be fine-tuned for many specific tasks. For example, models like GPT-3 (175B parameters), BERT (0.34B) or Indian efforts like BharatGen are foundation models. The key is *transfer learning*: once a model has "learned" language or vision patterns at scale, it serves as a starting point for various applications.

> *"Foundation models … form the base for a variety of AI applications."* – Telecom ET (May 2025) [11].

In other words, instead of building a new model for each use-case, developers can adapt (fine-tune) a foundation model. This is transformative for productivity. **However**, it also concentrates power: errors, biases or design choices in the base model propagate to many downstream systems. As the white paper notes, the design/training stage can "shape performance and risks across many downstream uses" (raising questions of sovereignty and safety).

- **Extra (not in document):** In practice, foundation models can inadvertently learn biases from their training data or "hallucinate" plausible but false information [12]. For instance, an IBM report warns that such models can generate toxic or misleading outputs. That's why governance is crucial – to catch and mitigate these risks.

Why the focus now? Generative AI (text, images, code, etc.) has exploded recently, showing the power of large models. India is racing to keep up: it has jumped to **#3** globally on Stanford's AI Vibrancy Index (2025) behind only the US and China [13] [10]. Investment-wise, India launched a massive **₹10,372 crore (~US$1.2B)** IndiaAI Mission in 2024 [14]. Its goal is to democratize AI compute, data and skills across the country. One pillar of this mission is precisely "Indigenous Foundation Models."

Building Indigenous Foundation Models

This section of the paper (Section 2) describes India's programs to **build and train AI models in India**. Key initiatives include:

- **Innovation Challenges / Teams:** IndiaAI issued calls for proposals and, through competitive selection, picked teams from industry and academia to create large models. By 2026, *twelve organizations/consortia* have been shortlisted (Sarvam AI, Soket AI, Gan AI, Gnani AI, Avataar, IIT Bombay's BharatGen consortium, GenLoop, Zenteiq, Intellihealth, Shodh AI, Fractal Analytics, Tech Mahindra) [1]. These teams get funding and computing resources to train multimodal AI (e.g., language+vision) on Indian data and languages. The models developed will be open-source and available to government and researchers, ensuring India-owned AI capabilities.

- **Compute Infrastructure:** A central theme is **democratizing compute power**. The paper notes initiatives like the India AI Compute Portal (public cloud platform) where startups and researchers access GPUs, storage and tools. For example, in March 2025 the government launched an AI Compute Portal with *10,000 GPUs initially and plans to add ~8,700 more* [3]. By early 2026, this expanded far beyond – India reports over **38,000 GPUs** now available across the country [2] [4]. These are offered at subsidized rates (₹65/hour mentioned) so even small labs or startups can train big models. In context, this is huge: a major challenge globally is that state-of-the-art AI requires clusters of high-end GPUs, often only affordable to tech giants. India's

approach is to pool government and private resources, e.g. partnerships with cloud firms, to build this capacity [15].

- *Extra:* Globally, AI compute is rapidly expanding – one Gartner report projects AI-related IT spending reaching **$2.5 trillion in 2026** [5]. India's public GPU project is part of that wave, but with a national twist on shared infrastructure.

- **Data Curation (AI Kosh):** Training foundation models needs vast, high-quality datasets. The government created **AI Kosh**, a data repository platform. AI Kosh aggregates curated datasets (text, images, audio, tabular data, etc.) across sectors (healthcare, finance, agriculture, etc.) and ensures they meet data privacy and quality standards. As of early 2025 it had over 300 datasets and 80+ models [6], and it has been growing explosively (internal reports say thousands by 2026). The white paper stresses AI Kosh's role: "by expanding access to large-scale, high-quality datasets, AIKosh supports foundation model training … reflecting India's linguistic and cultural diversity." Industry and academia are contributing data too (e.g. Soket AI's Indic language corpus).

- *Extra:* According to a PIB press release, AI Kosh had **5,500+ datasets and 251 models** by Dec 2025 [16] (covering 20 sectors). This shows rapid growth from just hundreds early on. For perspective, many international AI projects (like Meta's or Google's) scrape global data, but AI Kosh is intent on controlled, ethically-sourced Indian data. This aligns with India's data privacy laws (DPDP Act) and reduces reliance on foreign datasets [17].

- **Local Language Focus:** A recurring theme is linguistic diversity. India has 22 official languages and thousands of dialects. The initiatives emphasize models in Indic languages, and datasets with local context (health records, rural images, financial forms etc.). For instance, some of the selected teams explicitly work on multilingual models. The AI Kosh platform curates regional datasets so models learn Indian customs, units (e.g. understanding that a "bigha" area varies by state [18]), etc. This is critical – a foundation model trained only on English/Western data may poorly serve Indian users.

- **Talent and Research Hubs:** The mission also funds educational programs (supporting thousands of students and labs) to grow AI talent across India (see Section 4). For example, they mention establishing Data & AI labs in Tier-2/3 cities and sponsoring PhDs and fellowships. All of this ensures a pipeline of researchers and practitioners who can work on and with foundation models.

In summary, the "Building" section shows a multi-pronged effort: **fund R&D teams**, **provide massive compute**, **gather data**, and **train talent** – all to create "Made in India" foundation models. It's an infrastructure play plus innovation push.

Governing India's Foundation Models

Section 3 of the paper covers regulation and policy. The key messages:

- **India AI Governance Guidelines (2025):** India published a national AI governance framework centered on *seven principles* (called Sutras): Trustworthy, Human-Centric/People-First, Innovation over Restraint, Fairness & Equity, Accountability, Understandable by Design, Safety/Resilience/Sustainability [19]. These guidelines advocate a risk-based approach: stricter oversight for high-

impact AI applications, proportional to their potential harm. They emphasize transparency (e.g. making "transparency reports" about model risks and mitigation public) and human oversight. For example, developers are encouraged to document red-teaming results, impact assessments, biases found, etc. This mirrors global trends (like EU's AI Act or NIST's risk management) in tailoring rules to risk [20].

- **Digital Personal Data Protection (DPDP) Act 2023:** The white paper notes India's new data protection law as part of the landscape. DPDP imposes requirements on how personal data can be collected, processed, and shared – which directly affects AI training. Models must respect user privacy and data consent. The government highlights DPDP as giving "a robust legal framework for data privacy" that will govern AI as well [21]. This aligns with how the AI guidelines stress only using ethically-sourced data, not scraping private info without consent.

- **CERT-In and Safety Measures:** India's cybersecurity agency CERT-In has issued advisories on AI safety (e.g. against malicious use of generative AI) and launched a certification program (CSPAI) for AI security professionals. The white paper mentions these as part of "current landscape" of norms. There's also talk of creating an **AI Safety Institute** to centralize tech-policy research on AI security. These steps signal India's desire to stay current with global discussions on AI risks (cyber threats, deepfakes, fraud).

- **Regulatory Gaps:** Interestingly, the paper acknowledges that there is no AI-specific law yet; it relies on guidelines, existing laws (like IT Act, DPDP Act) and upcoming policies. This contrasts with some countries racing to pass AI Acts, but India is taking an "AI Governance by Principle" approach first, then building institutions and laws as needed.

- **Implications:** In short, India's policy stance is to **promote innovation but keep it accountable**. The Governance Guidelines explicitly mention a graded accountability along the AI value chain (developers, deployers, etc.) [19]. Essentially: encourage the development of powerful models *but* require checks on bias, misinformation, privacy, etc., especially for high-stakes use.

- *Extra:* This risk-based approach is in line with IBM's recommendation for policymakers (as noted in IBM's guide) to apply heavy regulation only where needed, to avoid stifling innovation [20]. It's also similar to frameworks like the EU's draft AI Act or Singapore's Model AI Governance, which use tiered rules.

AI for India's Priorities

Although Section 4's details are embedded in the white paper, the overall thrust is clear: tailor AI (especially generative AI) to India's socio-economic needs. The paper notes that generative AI should be applied to fields like healthcare, education, agriculture, vernacular services, etc. For example, they mention initiatives like AI solutions for medical imaging, farmers' advisory, local language education, or digital governance (even smarter Aadhaar services by a startup called Sarvam AI [22]).

Some highlights and extra context:

- **Agriculture:** Indian agriculture is a focus. (Extra: India has launched projects like *Bharat-VISTA*, a multilingual AI assistant for farmers that uses government data and ICAR knowledge to give advice [9]. This directly addresses farmers' needs in their language.) Given 600+ million in rural areas, AI in agri (predicting weather, disease in crops, optimizing inputs) is high priority.

- **Healthcare and Rural Services:** The pillars mention healthcare specifically (e.g., CoE in healthcare announced). AI models can help analyze X-rays, predict diseases, or manage health records in local languages. (Extra stat: globally, AI in healthcare is booming – projected $280 billion market by 2030, per some forecasts.) The white paper implies using foundation models for tasks like clinical diagnostics or medical chatbots tailored to Indian patient data.

- **Language and Education:** Making AI work in Hindi, Tamil, Telugu, etc., and preserving cultural context is key. (Extra: Google has efforts to bring AI to Indian languages and to connect farmers to education.) Education tech – personalized learning, tutoring bots in local tongues – is implied. The white paper also mentions "assistive technologies" in education sectors (from the press release [25]).

- **Governance and Public Services:** The government is looking at AI for better governance. As a concrete example, PIB reports that *Sarvam AI* is partnering with Aadhaar/UIDAI to use generative AI for smarter identity verification and secure document understanding [22]. This suggests foundation models can streamline citizen services (e.g. automatically answering queries, fraud detection in IDs, etc.).

- **Climate and Infrastructure:** While not always headline-catching, AI can optimize energy use, predict disasters, and plan smart cities. The mission's pillars mention climate and sustainable cities too.

In essence, the paper wants to tie foundation models into national development goals. These large models are seen as general-purpose "engines" that can be fine-tuned for any domain, so the emphasis is on building them now so they can be applied across critical sectors later.

Key Statistics and External Context (Extra Information)

- **IndiaAI Mission Scale:** As of early 2026, IndiaAI Mission has *a Rs10,372 Cr (~$1.3B) budget* over 5 years [23]. It's organized into *seven pillars* (Compute, Datasets/AI Kosh, Applications, Foundation Models, FutureSkills, Startup Financing, Safe & Trusted AI) [24] [25]. Notably, India has onboarded ~38,000 GPUs (exceeding initial 10k target) under the Compute pillar [4], making it one of the largest shared AI supercomputing grids globally.

- **Foundation Model Projects:** India received *500+ proposals* for building large AI models; only 12 were selected [26]. This suggests high interest. Selected teams include established companies (Tech Mahindra, Fractal) and startups. By May 2025, 3 startup teams (Soket, Gan, Gnani) were chosen alongside Sarvam (earlier) [27]. This parallels a similar program in the US/UK to foster local LLMs.

- **Datasets Growth:** AI Kosh saw explosive growth. In May 2025, there were 367 datasets [28]; by the white paper's writing (Feb 2026) over *10,000 datasets* are hosted (as the PDF notes). By end-2025, official stats cite 5,500+ datasets, 251 models [16]. This means thousands of new datasets were added in months – a massive indexing push. (Extra: In a global context, Google's public *Language Models* often reuse open data, but India's approach is to build its own curated corpora. The quantity suggests big data efforts by government and community.)

- **Startups and Ecosystem:** The press releases mention 30 government-approved AI applications (by 2026) and 12 FM teams [29]. Private players are engaged: Indian startups like Gnani AI, Soket AI, Gan AI, Fractal Analytics, Avataar AI etc. Many funded projects have ties to IITs or BITS. There's also an *AI Startups Global* program sending 10 Indian AI startups to France to scale up (March 2025) [30].

- **Global & Economic Context:**

- *AI spending:* Worldwide AI spending is skyrocketing – Gartner predicts \$2.52 trillion in 2026 (a 44% jump over 2025) [5]. A significant portion of that is infrastructure (~\$1.37T on AI servers/ platforms). India's Rs10kCr (~\$1.25B) is modest globally, but targeted.
- *AI Vibrancy Index:* Stanford's Global AI Vibrancy (2025) ranks India 3rd [10]. This index considers investment, talent, number of technical professionals, etc. (Extra: India's strength comes from its IT talent pool and growing startup scene; however, it still lags US/China in proprietary AI breakthroughs.)

- *GDP and AI:* Indian policymakers see AI as a growth engine. (Extra: McKinsey estimates that AI could add up to 1.1% per year to India's GDP growth by 2030 via productivity gains.)

- **Emerging Policies:** India's AI Guidelines (2022 draft, finalized 2025) and DPDP Act make a clear policy stance. The white paper echoes these: emphasizing "transparency reports," risk categories, and institutional bodies (AI Governance Group, AI Safety Institute) [31]. This alignment is in step with global norms – e.g. EU's draft law and NIST's framework also advocate risk-based measures. (Extra: By adopting "Innovation over Restraint," India signals it wants a lighter regulatory touch in low-risk areas, unlike blanket bans in some regions.)

- **International Partnerships:** Not detailed in the doc, but India is collaborating internationally. (Extra: For example, India participated in G7's AI discussions; there are MoUs with countries on AI research; Indian teams often use open-source architectures from OpenAI, Google, etc. The PSA office paper does not delve into geopolitics, focusing on domestic moves.)

Conclusion & Talking Points

In summary, the foundation model white paper tells us that India is *actively building* its own generative AI ecosystem. Rather than relying entirely on imported AI (like commercial GPT or Stable Diffusion), India wants to train models on its own data, languages and needs. The rationale is clear: sovereignty, customization, and economic opportunity.

Key takeaways are: - **Infrastructure:** Massive investment in GPUs and data means India's labs can actually train big models. - **Local Talent & Teams:** By funding Indian researchers and startups to build models, the country hopes to create homegrown AI champions. - **Policy & Safety:** Simultaneously, India is putting in place rules (guidelines, laws) so that as AI grows, it's aligned with Indian values (privacy, fairness, security). - **National Goals:** The end-use of these models is envisioned to be wide – from smarter public services to boosting farming, education, healthcare, etc., in an inclusive way.

Talking Points / Q&A:

- **Q: What is a foundation model, in simple terms?**
  A: It's a massive AI model trained on broad data (text, images, etc.) that serves as a starting point for many applications. Think of it like a generic brain that you can fine-tune to specific tasks. For

example, GPT-3 is a foundation model: pre-trained on internet text, it can generate essays, answer questions, or even write code once customized [11] [32].

- **Q: Why is India focusing on building its own foundation models?**
  A: To ensure AI reflects Indian languages, culture and priorities. If only global models (trained mostly on English/US data) are used, they may not work well for Indians. India's strategy is to create sovereign models so it controls the tech and jobs. Plus, building these models grows local expertise and keeps economic value in India.

- **Q: What are India's main initiatives in this area?**
  A: Under the IndiaAI Mission (2024-29), there are several pillars: one is "Foundation Models" (teams building new models), one is "Compute" (GPUs and cloud), one is "Datasets" (AI Kosh), plus others like applications development and skills. Specifically, the government issued calls for proposals and now backs 12 teams to make foundation models [1]. They also set up the AI Compute Portal with tens of thousands of GPUs on demand [3] [4], and launched AI Kosh to centralize data. In practice, this means you, a startup or researcher, can apply for compute credits or use the data library to train a model.

- **Q: What is AI Kosh?**
  A: It's a platform (like a catalog) of Indian datasets, tools, and even some pre-trained models. Developers can access thousands of datasets (text corpora, government data, images, etc.) for free or at low cost, with permission controls. The idea is to share non-sensitive data widely so that models can be trained effectively. It also helps ensure data used are compliant with privacy laws.

- **Q: What about regulations and safety?**
  A: India has released voluntary guidelines (with 7 principles like "Trust" and "Accountability") and a Data Protection Act (2023). There are also new bodies (AI Governance Group, Safety Institute) being set up. In essence, the government says: "Build AI, but do it responsibly." For example, if you create a high-risk AI (like one that decides medical outcomes), you should follow stricter checks (document bias testing, have human oversight, etc.).

- **Q: How does this fit in the global picture?**
  A: India is playing catch-up but moving fast. Globally, tech giants in US/China dominate model development (OpenAI, Google, Baidu, etc.). Countries like the US, UK, China, EU all have similar initiatives for homegrown AI. India's unique advantage is its large pool of software engineers and growing startup scene. It's also now a top-3 AI nation in terms of activity [10]. Big tech companies are investing heavily in India's AI ecosystem (Amazon $35B, Microsoft $17.5B, Google $15B commitments [33]). The global AI trend is massive spending (AI infrastructure grew 17% in 2026 [34]). India's share is still small but focused.

- **Q: Should I mention any statistics?**
  A: Definitely mention GPU and team numbers (e.g., "~38,000 GPUs made available [2] and 12 funded teams for models [1]"). Also mention the budget (₹10,372 Cr) and students supported (thousands of UG/PG/PhD) [2]. It shows scale. And quotes from press info like "India ranks #1 in AI skill penetration" can emphasize progress [3]. Finally, global context: "Worldwide AI spending will reach ~$2.5 trillion by 2026 [5]" underscores why AI is critical.

Extra: The white paper itself doesn't list specific impact metrics (like accuracy improvements or new product launches), since it's a policy/strategy document. But you can add that similar initiatives

internationally have accelerated AI adoption. For example, when the US launched its AI ecosystem funding, many research breakthroughs (like AlphaFold in biotech) followed. India hopes to trigger similar innovations by providing these building blocks.

Keep in mind, your discussion should convey that you **understand the document's content**: it's about India's plan for big AI models and how to govern them. Use the facts above to speak confidently about what's happening (e.g., "The paper emphasizes that choices in these large models affect many applications, so India is stressing responsible design."). You can sprinkle in the extra stats and context we've gathered to show you've done broader research. Good luck for your discussion!

---

1 2 7 19 21 23 29 31 Press Release: Press Information Bureau
https://www.pib.gov.in/PressReleseDetailm.aspx?PRID=2227612®=3&lang=2

3 6 17 30 Press Release:Press Information Bureau
https://www.pib.gov.in/PressReleasePage.aspx?PRID=2108961®=3&lang=2

4 10 16 22 24 25 26 Press Release:Press Information Bureau
https://www.pib.gov.in/PressReleasePage.aspx?PRID=2209737®=3&lang=2

5 34 Gartner Says Worldwide AI Spending Will Total $2.5 Trillion in 2026
https://www.gartner.com/en/newsroom/press-releases/2026-1-15-gartner-says-worldwide-ai-spending-will-total-2-point-5-trillion-dollars-in-2026

8 12 20 A Policymaker's Guide to Foundation Models
https://newsroom.ibm.com/Whitepaper-A-Policymakers-Guide-to-Foundation-Models

9 13 18 33 AI in India: Small Language Models for big Impact in farming
https://agfundernews.com/ai-in-india-the-worlds-ai-back-office-is-betting-on-small-language-models-to-bring-big-impact-to-smallholder-farming

11 27 28 Gnani AI: India Expands AI Initiative with New Startups and Enhanced GPU Capacity, ETTelecom
https://telecom.economictimes.indiatimes.com/news/internet/india-expands-ai-initiative-with-new-startups-and-enhanced-gpu-capacity/121528404

14 15 IndiaAI Mission: Democratizing Access to AI Compute
https://www.ibef.org/blogs/indiaai-mission-democratizing-access-to-ai-compute

32 What are foundation models for AI?
https://www.redhat.com/en/topics/ai/what-are-foundation-models