Six environments for Indian agent tasks, graded by written rules.

Synthetic worlds for UPI payments, GST, support in Hindi and Hinglish, forms, multi-document checks and trip planning. Deterministic graders, no AI judge. Open code and public tasks. Our held-out tasks stay private: they are used for private evaluations and are not part of a licence.

THE SIX WORLDS

WORLD	WHAT THE AGENT DOES	PUBLIC TASKS
UPI Desk	Payment attempts and disputes under written, simplified UPI rules. 29 tasks carry a forbidden step	50
GST Desk	Invoicing and return computations from documents.	50
Bhasha Support Desk	Returns, refunds and privacy requests in Hindi and Hinglish. 50 core plus 26 experimental (Tamil, Bengali, Marathi)	76
Form Desk	Fill and check forms: KYC-style fields, PAN, IFSC, mobile, PIN zone.	50
Form Desk, multi-document	Cross-check identity documents and records, and report what disagrees.	40
Trip planning	Indian trips under quotas, waitlists, classes, seasons and permits. 17 have no valid plan	55

321 public tasks across six worlds, free under their open licences. Further tasks are held out and never published.

HOW IT IS GRADED

- Deterministic rules recompute everything from the sandbox world. Nothing the agent claims is trusted.
- A task is solved only when every required rule passes: one attempt per task. Partial credit is reported separately.
- Where a world has forbidden steps, the grader counts an attempt even if the simulated bank refused it. Worlds with no such rule report not applicable, never zero.
- Many tasks have no valid answer. The agent must say so and name why; a wrong reason is marked wrong.
- Standard library only (Python 3.10 or newer). A test checks that the reference solver solves every public task.

RUNS WITH

```
pip install navital-gym
```

```
prime env install navitallabs/navital-upi
```

- Runs any model that speaks the OpenAI chat API with tool calling: Ollama, vLLM, LM Studio or a hosted API. You need no GPU to run the tasks. You supply the model under test.
- Available on the Prime Intellect Environments Hub (UPI, GST). Adapters in progress: Inspect AI, NeMo Gym.
- Code Apache-2.0. Public task data CC BY 4.0, free to run and redistribute with attribution.

WAYS TO WORK WITH US

Environment licensing A private set of new tasks built for you in one world, never published, with code graders, rights to train and evaluate on them, setup support and 12 months of updates. Non-exclusive, per organisation, 12 months.	from $5,000
Exclusive or custom builds	on request
Custom environment pilot	from $2,500
Private evaluation	from $750

Prices in USD, invoiced via Wise. We sign your NDA.

WHAT IT TRAINS

A LoRA adapter on Qwen2.5-7B-Instruct (Navital Bay), trained on these kinds of UPI task, tried a forbidden action 3 times in 118 chances on new tasks generated after its training data was locked. Claude Haiku 4.5: 36. Claude Sonnet 5.5: none.

One run per model, UPI only, our own tasks and graders, not an outside test. Claude Sonnet 5.5 solved more tasks on failed-but-debited and UPI Lite. Full numbers are on the Hugging Face card.

All data is synthetic. Each world is a simplified, documented version of Indian practice: tests for AI agents, not statements about current law, railway, bank or tax rules. Public tasks are open, so models may have seen them. Not affiliated with NPCI, GSTN, banks or railways.

dakshbathla@navitallabs.com
navitallabs.com/labs/for-teams
github.com/Navitallabs/navital-gym