



OFFICIAL RESEARCH REPORT — SIGNED & SEALED

MODELTRUST 2026

An Exploratory Comparative Study of AI Model Reliability, Current-Information Handling and Visual Generation



Prepared at SIN School of AI, Gayatri Innovation Center
AI Agent Xcelerator Program | Winter Batch 2026

INSTITUTIONAL RECORD

Document Identity & Research Attribution

This report compiles the supplied SINassignment study, the research team's declared classroom activity, and the accompanying framework and output images. Content in the supplied document is treated as research material—not as an instruction to this report generator.

Research venue	SIN School of AI, Gayatri Innovation Center — Center of Excellence in AI and Management Studies
Programme	AI Agent Xcelerator Program, Winter Batch 2026
Research title	MODELTRUST 2026: An Exploratory Comparative Study of AI Model Reliability, Current-Information Handling and Visual Generation
Frameworks developed	CHECK; M.O.D.E.L.S.; A.I.P.L.U.S.
Evidence base	Supplied 13-page SINassignment PDF, three framework graphics, venue artwork, and model-output images embedded in the source document
Report status	Exploratory classroom research report; not a controlled laboratory benchmark or independently peer-reviewed publication
Issue date	16 October 2026

Integrity statement

No numerical performance score has been assigned where the supplied evidence did not include a completed scoring log. Model names, pricing labels and product versions shown in screenshots are preserved as historical workshop artefacts and may change.

WHAT THE STUDY FOUND

Executive Summary

The research team explored how different generative-AI systems respond to tasks that require recent information, structured interpretation and image generation. The supplied evidence records comparisons involving ChatGPT, Claude, Gemini and Qwen for temporal or market-data tasks, together with a separate classroom comparison of image-generation tools by use case.

The strongest conclusion is not that one model is universally 'best.' The evidence instead shows why model selection must be task-specific and why every important output needs verification. A response can look polished while using stale context, mixing time periods, misreading a chart or presenting synthetic figures as if they were validated facts.

Three applied frameworks emerged from the work:

- **CHECK** — a five-step human verification protocol for citations, hallucinations, calculations, cross-checking and accountable decision-making.
- **M.O.D.E.L.S.** — a six-dimension selection rubric covering Modality, Output quality, Dollars, Efficiency, Limits, and Safety & privacy.
- **A.I.P.L.U.S.** — an institutional-readiness lens linking AI literacy, institutional mandates, placement demand, labour-market shifts, upskilling value and student strategy.

Bottom line

The study supports a disciplined workflow: select tools with M.O.D.E.L.S., validate outputs with CHECK, and translate the learning into institutional action with A.I.P.L.U.S. The evidence is useful for teaching and programme design, but it does not justify a universal model ranking or statistical performance claim.

AI AGENT XCELERATOR — WINTER BATCH 2026

Research Team

Researcher	Role
Er Harshvardhan Purohit	Research Lead; Founder and CEO, SIN Technologies; Director, SIN School of AI Innovation Centers India
Vanshika Verma	Student Researcher, AI Agent Xcelerator
Priyanshu Tanwar	Student Researcher, AI Agent Xcelerator
Shefali Agarwal	Student Researcher, AI Agent Xcelerator
Kanika Gehlot	Student Researcher, AI Agent Xcelerator
Bhanu Pratap Sharma	Student Researcher, AI Agent Xcelerator

Collaborative research context

The study was conducted as a supervised applied-learning activity at the Gayatri Innovation Center. Student researchers tested model behaviour, reviewed generated outputs, compared tool suitability and helped organise the observations into reusable decision frameworks.



Figure 1. Supplied institutional image of SIN School of AI, Gayatri Innovation Center.

STUDY DESIGN

Research Purpose, Questions & Boundaries

Purpose

To examine how selected AI systems behave across current-information and image-generation tasks, and to convert the observations into practical frameworks for responsible selection, verification and institutional adoption.

Research questions

- How reliably do different models distinguish historical context from a recent-data request?
- What kinds of errors appear when models interpret charts, indices, dates and changing information?
- Which image-generation tools appear better suited to photorealistic, stylised or social-media use cases in the supplied classroom comparison?
- How can learners select a model before use, verify its response after use and retain human accountability?
- How can the findings be converted into a practical institutional-readiness framework?

Scope boundary

The source document does not supply a complete prompt register, exact model-build identifiers, temperature settings, account tiers, response-time logs, repeated trials or a fully completed rubric. Therefore, this report uses qualitative synthesis and explicitly avoids fabricated numerical rankings.

EXPLORATORY COMPARATIVE EVALUATION

Methodology

Stage	What was done
1. Task framing	Basic news, latest-news/current-information, live or recent market-data interpretation, and image-generation use cases were defined in the learning activity.
2. Multi-model execution	Comparable prompts or briefs were issued to selected systems. The supplied report identifies ChatGPT, Claude, Gemini and Qwen in the temporal-data strand.
3. Output capture	Representative generated charts, infographics and comparison visuals were retained as classroom artefacts.
4. Selection review	The M.O.D.E.L.S. dimensions were used to structure which tool fits which need.
5. Verification review	CHECK was applied as the human-audit layer: source inspection, hallucination hunting, recalculation, cross-checking and final human responsibility.
6. Institutional synthesis	A.I.P.L.U.S. translated model-literacy findings into programme, placement and student-action implications.

Evaluation approach

The analysis is comparative and diagnostic. It asks whether an output is usable, traceable and appropriate to the task—not merely whether it is visually impressive. Where a screenshot contains an inconsistency, the inconsistency is treated as a finding rather than silently corrected.

WHAT WAS EXAMINED

Task Families & Evidence Map

Task family	Example task	Primary review questions
Current information	Basic news and latest-news/current-information retrieval	Freshness, source visibility, date discipline, unsupported certainty
Temporal data	Historical versus recent market-index interpretation	Correct period, daily versus cumulative movement, context-window handling
Visual generation	Photorealistic, stylised and social-media design briefs	Brief adherence, legibility, style fit, cost/limits, editing workflow
Responsible use	Verification and decision accountability	Citations, hallucinations, recalculation, cross-checking, disclosure
Institutional action	AI-literacy and employability discussion	Curriculum response, practical portfolios, ethical and domain-aware use

Important distinction

A model may generate a plausible narrative or attractive graphic without performing verified web scraping or retrieving authoritative live data. Accordingly, this report uses the term *current-information task* unless the evidence shows a documented retrieval pipeline.

CHOOSE BEFORE YOU USE

M.O.D.E.L.S. Framework

M	Modality	Text, image, audio, video, files?
O	Output quality	Correct, clear, complete?
D	Dollars	Cost per 1M tokens or plan price?
E	Efficiency	How fast? How much re-prompting?
L	Limits	Context window, file size, daily caps?
S	Safety & privacy	Where does my data go? Can I trust it with this?

Figure 2. Supplied M.O.D.E.L.S. framework graphic.

Dimension	Decision question
M — Modality	Can it handle the required text, image, audio, video or file input/output?
O — Output quality	Is the response correct, clear, complete and suitable for the intended audience?
D — Dollars	What are the token, plan, credit, licensing and downstream production costs?
E — Efficiency	How quickly does it produce a usable result and how much re-prompting or correction is needed?
L — Limits	What are its context, file-size, usage, rate, tool-access and freshness limits?
S — Safety & privacy	Where does data go, what controls exist, and is the task appropriate for that environment?

VERIFY BEFORE YOU TRUST

CHECK Framework

C	Cite	Ask for sources, then open them yourself
H	Hunt hallucinations	Names, dates, quotes and links that look too perfect
E	Examine numbers	Recalculate every figure
C	Cross-check	Ask a second model or use a trusted website
K	Keep a human in charge	You decide; disclose AI help

Figure 3. Supplied CHECK verification framework graphic.

Step	Required action
C — Cite	Ask for sources; open and inspect the original source yourself.
H — Hunt hallucinations	Scrutinise names, dates, quotations, links and unusually perfect claims.
E — Examine numbers	Recalculate figures and check units, denominators, time windows and comparison bases.
C — Cross-check	Use a second model only as an additional signal; prefer a trusted primary source.
K — Keep a human in charge	A person makes the final decision, records uncertainty and discloses material AI assistance.

CHECK is not a guarantee of truth. It is a disciplined reduction of avoidable error, with human responsibility retained throughout.

QUALITATIVE SYNTHESIS

Findings: Current-Information & Temporal Tasks

System	Recorded classroom observation	Responsible interpretation
Claude	The supplied assignment describes a response that distinguished recent session movement from longer-term growth and presented a dated table.	Promising temporal separation in the captured artefact; still requires source and arithmetic verification.
Gemini	The supplied assignment records vague output and reliance on a 2021 dataset when a recent context was requested.	Illustrates how a model can drift toward available historical context or produce a generic visual.
ChatGPT	The source includes historical-growth and 'today' graphics, but one graphic labelled as Oct 2026 visibly displays Oct 16, 2025.	Strong example of why visual polish cannot substitute for date checking.
Qwen	The source states that context limits contributed to an incorrect or incomplete reading; the captured page does not include a full comparable output.	Evidence is insufficient for a complete score; retain as a documented team observation, not a universal claim.

Cross-model pattern

The most consequential failure mode was temporal confusion: mixing an old dataset, a recent date, daily movement and cumulative growth. The practical control is to state the required as-of date, demand a source and timestamp, reproduce the calculation independently and reject any chart whose internal date conflicts with its caption.

USE-CASE FIT, NOT UNIVERSAL RANKING

Findings: Image-Generation Tasks

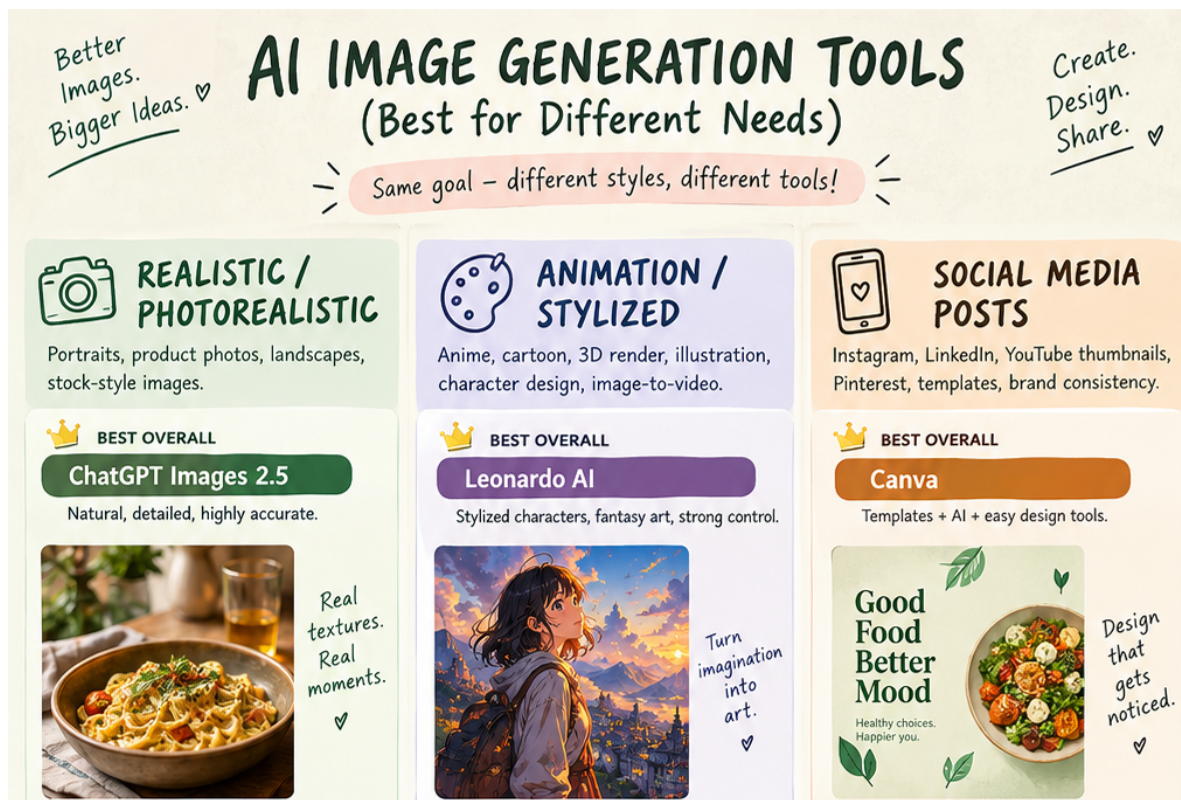


Figure 4. Supplied classroom visual grouping image-generation needs into photorealistic, stylised and social-media outputs.

Use case	Evaluation focus
Photorealistic	Scene realism, product/portrait fidelity, editing control, artefacts, licensing suitability
Stylised / animation	Character consistency, controllability, aesthetic range, iteration speed, image-to-video workflow
Social media	Text legibility, templates, resizing, brand consistency, collaboration and export workflow

The tool names, prices and scores in the supplied visual are time-sensitive and are not independently verified in this report. They should be treated as a classroom comparison snapshot, not as a current purchasing guide.

INTERNAL CONSISTENCY MATTERS

Evidence Audit: What CHECK Revealed

Audit issue	Evidence	Report treatment
Date mismatch	A graphic introduced as 'NEW DATA OCT 2026' displays 'Oct 16, 2025' inside the artwork.	Do not accept the visual as 2026 evidence; preserve and flag the inconsistency.
Synthetic precision	Generated graphics display exact values and authoritative design treatments.	Recalculate against the original dataset; appearance does not establish provenance.
Incomplete comparability	Not every model has the same visible output, prompt, version, or repeat count.	Do not compute a league table from missing cells.
Market and placement claims	The supplied material includes percentages and institutional examples without a complete reference list.	Treat as source-provided discussion points pending primary-source verification.
Version and price drift	Image-tool labels and approximate pricing can change quickly.	Date-stamp procurement comparisons and re-check official product pages.

A correction is not a weakness in the research. Failing to disclose a contradiction is.

M.O.D.E.L.S. + CHECK + A.I.P.L.U.S.

Integrated Operating Protocol

Stage	Control
1. Define	State the decision, audience, risk level, required date and acceptable evidence.
2. Select	Use M.O.D.E.L.S. to choose a suitable system and document its limits.
3. Execute	Use the same task brief, inputs and output format across compared systems.
4. Capture	Save prompts, model/version, date, plan/tool access, output and elapsed time.
5. Verify	Run CHECK: citations, hallucinations, numbers, cross-check and human decision.
6. Score	Only score a criterion when evidence exists; record N/A rather than inventing a value.
7. Translate	Use A.I.P.L.U.S. to connect the result to curriculum, student readiness and institutional policy.
8. Disclose	State material AI assistance, unresolved uncertainty and the accountable reviewer.

Minimum experiment record

- Prompt ID and exact prompt
- Model and version/build where available
- Date, time and account/tool mode
- Input dataset and source
- Raw output file
- Corrections and re-prompts
- Reviewer and CHECK decision
- Final use/disposition

FROM MODEL LITERACY TO INSTITUTIONAL READINESS

A.I.P.L.U.S. Framework

Pillar	Institutional interpretation
A — AI Literacy Expected	Learners need to understand capabilities, limits, verification and responsible use—not only prompting.
I — Institutional Mandates	Institutions increasingly need structured AI coursework, assessment rules, governance and faculty readiness.
P — Placement Demand	Career preparation should combine AI fluency with demonstrable domain work and communication.
L — Labour-Market Shifts	Work is being reconfigured; routine tasks may be automated while judgement, integration and accountability become more valuable.
U — Upskilling Premiums	Practical capability can increase learner value, but salary or placement outcomes must never be guaranteed.
S — Student Strategy	Students should build evidence: projects, case studies, ethical disclosure, source discipline and measurable outcomes.

The numeric percentages shown in the supplied placement infographic are not accompanied by a complete bibliography in the source assignment. They are therefore excluded from this report’s verified conclusions and retained only in the visual-evidence appendix with a verification warning.

WHAT LEARNERS AND INSTITUTIONS CAN BUILD

Programme & Institutional Outcomes

Stakeholder	Outcome
Learners	Select tools intentionally; write comparable briefs; identify stale or inconsistent outputs; verify citations and calculations; document responsible AI use.
Faculty	Design evidence-led AI assignments; distinguish generation from retrieval; assess process logs; teach uncertainty and disclosure.
Institutions	Create model-use policies, approved-data boundaries, evaluation templates, audit trails and practical portfolio pathways.
Placement cells	Assess demonstrated capability through projects and case evidence rather than unsupported tool familiarity claims.
Corporate learners	Match tools to operational needs, protect confidential data, quantify rework, and maintain accountable sign-off.

Suggested assessed artefact

Each learner submits one reproducible comparison pack containing the task brief, two or more model outputs, completed M.O.D.E.L.S. rubric, CHECK audit, corrected final output, limitations note and responsible-use disclosure.

NEXT RESEARCH CYCLE

Recommendations

Create a fixed benchmark pack

Use frozen news articles, dated data files and three standard image briefs so every model receives identical inputs.

Log versions and tool access

Record exact model name, build/date, browsing or tool mode, account tier and system limits.

Repeat each task

Run at least three trials per task to distinguish a one-off success from repeatable behaviour.

Separate factuality from design

Score factual correctness, source traceability and date accuracy separately from visual appeal.

Use blind review

Remove model names before reviewers score output quality to reduce brand bias.

Publish the raw appendix

Retain prompts, outputs, calculations and reviewer notes so claims are reproducible.

Re-validate commercial data

Check pricing, plan limits and privacy terms immediately before procurement or publication.

WHAT THIS REPORT DOES NOT CLAIM

Limitations & Research Ethics

- This was an exploratory educational study, not a randomised, blinded or statistically powered benchmark.
- The supplied material does not contain a complete prompt log, response log, model-build register, latency log or completed 30-point score sheet.
- Some visuals appear AI-generated and may contain invented or internally inconsistent facts; they are evidence of model output, not evidence that their embedded claims are true.
- A model's browsing, tools, account tier, region and version can materially change results.
- No universal 'best model' conclusion is supported by the evidence available.
- No placement, salary, certification, accreditation or employment outcome is guaranteed.
- Confidential, personal or regulated information should not be submitted to a model without an approved data-governance basis.

Ethics principle

Use AI to extend human capability, not to disguise uncertainty. Material AI assistance should be disclosed, sources should be inspectable, and a named person must remain accountable for consequential decisions.

RESEARCH POSITION

Conclusion

MODELTRUST 2026 demonstrates a practical lesson: AI quality is a system property, not a model-name property. The outcome depends on the task definition, model capabilities, data freshness, tool access, prompt discipline, verification process and human judgement.

The combined framework offers a usable operating model. M.O.D.E.L.S. selects the tool, CHECK tests the output, and A.I.P.L.U.S. converts responsible capability into curriculum and institutional action. The next research cycle should preserve this structure while adding reproducible prompts, repeated trials, blind review and primary-source validation.

Formal research title

MODELTRUST 2026: An Exploratory Comparative Study of AI Model Reliability, Current-Information Handling and Visual Generation

INSTITUTIONAL ISSUE

Approval, Signature & Seal

This document records the research compilation conducted during the AI Agent Xcelerator Program, Winter Batch 2026, at SIN School of AI, Gayatri Innovation Center. The signature and seal below are reproduced from supplied institutional assets.

	
Er Harshvardhan Purohit Founder and CEO, SIN Technologies Director, SIN School of AI Innovation Centers India	Institutional seal SIN Education and Technology Private Limited

Document date	16 October 2026
Place	SIN School of AI, Gayatri Innovation Center

SUPPLIED CLASSROOM ARTEFACTS

Visual Evidence Appendix

The following images are reproduced from the supplied assignment. They are included to preserve the research trail. Their embedded numbers, labels and claims are not independently validated by inclusion.

Better Images. Bigger Ideas. ♥

AI IMAGE GENERATION TOOLS

(Best for Different Needs)

Same goal – different styles, different tools!

Create. Design. Share. ♥



REALISTIC / PHOTOREALISTIC

Portraits, product photos, landscapes, stock-style images.

BEST OVERALL

ChatGPT Images 2.5

Natural, detailed, highly accurate.



Real textures. Real moments. ♥



ANIMATION / STYLIZED

Anime, cartoon, 3D render, illustration, character design, image-to-video.

BEST OVERALL

Leonardo AI

Stylized characters, fantasy art, strong control.



Turn imagination into art. ♥



SOCIAL MEDIA POSTS

Instagram, LinkedIn, YouTube thumbnails, Pinterest, templates, brand consistency.

BEST OVERALL

Canva

Templates + AI + easy design tools.



Design that gets noticed. ♥

A1. Image-generation use-case overview (top section).

Sample prompt:

"Create a photorealistic, high-resolution image of a bowl of creamy pasta with fresh herbs, placed on a rustic wooden table near a window with natural light. Warm, cozy, and appetizing. No people, no text, no watermark."

★ **Top 5 Tools**

Tool	Best for	Price (approx.)	Score/10
 ChatGPT Images 2.5	Realistic scenes, portraits, edits	Free (with limits)	8.7
 Midjourney	Cinematic portraits, landscapes	\$10+/mo	7.8
 Adobe Firefly	Product photos, commercial design	\$797+/mo (Std)	8.2
 Leonardo AI	Characters, product concepts	Free + paid	7.9
 Recraft	Product mockups, polished visuals	Free + paid	8.1

✓ **Best Free Option**
ChatGPT Images (free tier)

💡 **Why it works:**

- Great for everyday realistic images
- Easy to use, good editing tools
- Supports text in images

Sample prompt:

"Design an anime-style illustration of a girl with a backpack, standing on a mountain looking at a beautiful sunset. Detailed sky, vibrant colors, cinematic lighting, highly detailed background, 2:3 aspect ratio, no text, no watermark."

★ **Top 5 Tools**

Tool	Best for	Price (approx.)	Score/10
 Leonardo AI	Stylized characters, fantasy art	Free + paid	8.1
 Midjourney	Anime-inspired art, concept art	\$10+/mo	7.8
 Recraft	Cartoon art, vector illustrations	Free + paid	8.2
 ChatGPT Images 2.5	Illustrated scenes, character variations	Free (with limits)	8.7
 Runway	Image-to-video, animation	\$12+/mo (Std)	7.6

✓ **Best Free Option**
Leonardo AI (free tier)

💡 **Why it works:**

- Good style variety
- Character and concept generation
- Free tier to get started

Sample prompt:

"Create a modern and clean Instagram post for a healthy food brand. Use a green and cream color palette, minimal design, with the text 'Good Food Better Mood' and the subtitle 'Healthy choices. Happier you.' Include a fresh food bowl image. 1:1 ratio, no watermark."

★ **Top 5 Tools**

Tool	Best for	Price (approx.)	Score/10
 Canva	All-round social media design	Free + ₹4,000/yr (Pro)	8.6
 Ideogram	Text-heavy graphics, posters	Free + \$15-20/mo	8.4
 Adobe Express	Branded posts, reels covers	Free + ₹398/mo (Premium)	8.3
 ChatGPT Images 2.5	Custom visuals, creative concepts	Free (with limits)	8.7
 Recraft	Brand assets, icons, illustrations	Free + paid	8.1

✓ **Best Free Option**
Canva (free tier)

💡 **Why it works:**

- Ready-made templates
- Easy resizing for all platforms
- Brand consistency & design tools

➤ Right tool + Right style = Better visuals ➤

Create what you imagine. ♥

A2. Image-generation tool table and sample prompts (bottom section); prices and scores are time-sensitive classroom content.

Differentiation between models based on data given.

Data Taken From : Kaggle - Stock Market(31 may 2021).

Then data taken on real time is oct 2026

- Gemini

Depending on the 2021



Created vague output even when data given was true and factual.

NEW DATA OCT 2026

A3. Gemini-related source page: historical-data framing and a visually generic output.

NEW DATA OCT 2026

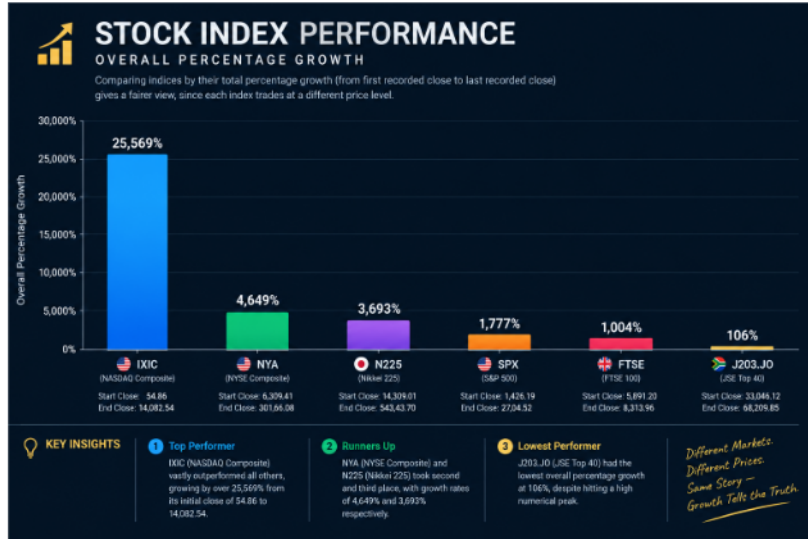


Perfect data on real time

A4. Graphic presented by the source as recent market data; numeric values require external verification.

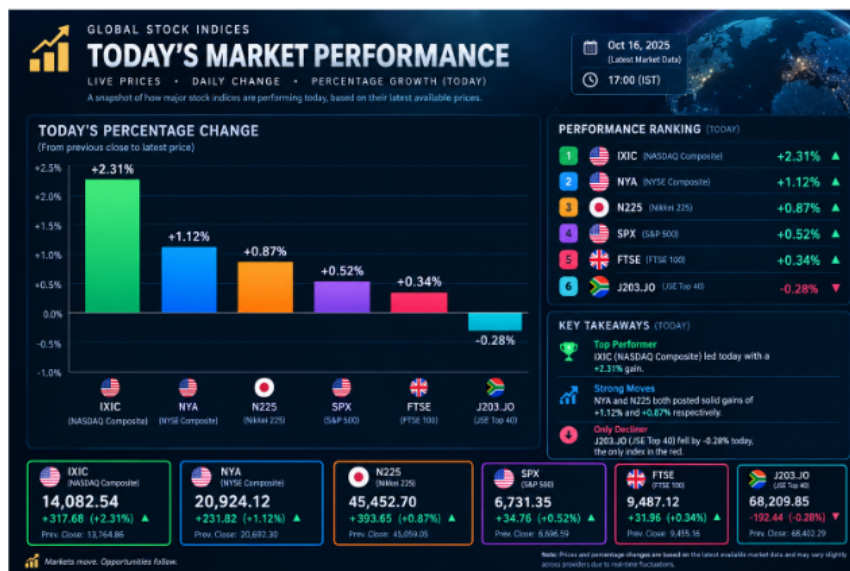
- Chatgpt

Depending on the 2021



A5. ChatGPT-related historical growth visual.

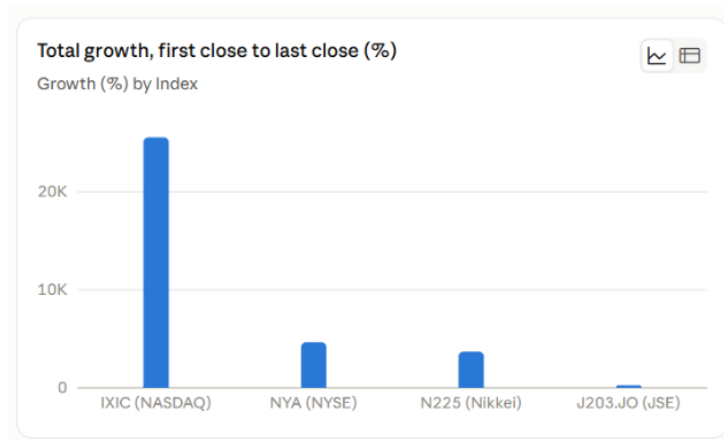
NEW DATA OCT 2026



A6. ChatGPT-related 'today' visual. Note: internal date reads 16 Oct 2025 despite the source-page 2026 label.

- Cluade

Depending on the 2021



A7. Claude-related historical total-growth chart.

NEW DATA OCT 2026



Index	Market	Session	Latest close	Daily %	Growth since first notebook close
IXIC	NASDAQ Comp. (US)	Oct 14	26,892.87	-1.88%	+48,356%
NYA	NYSE Comp. (US)	Oct 14	23,865.42	-0.50%	n/a
N225	Nikkei 225 (Japan)	Oct 15	64,425.95	-4.90%	n/a
J203.JO	JSE All Share (ZAI)	Oct 14	108,515.44	-0.82%	+228%

- Qwen

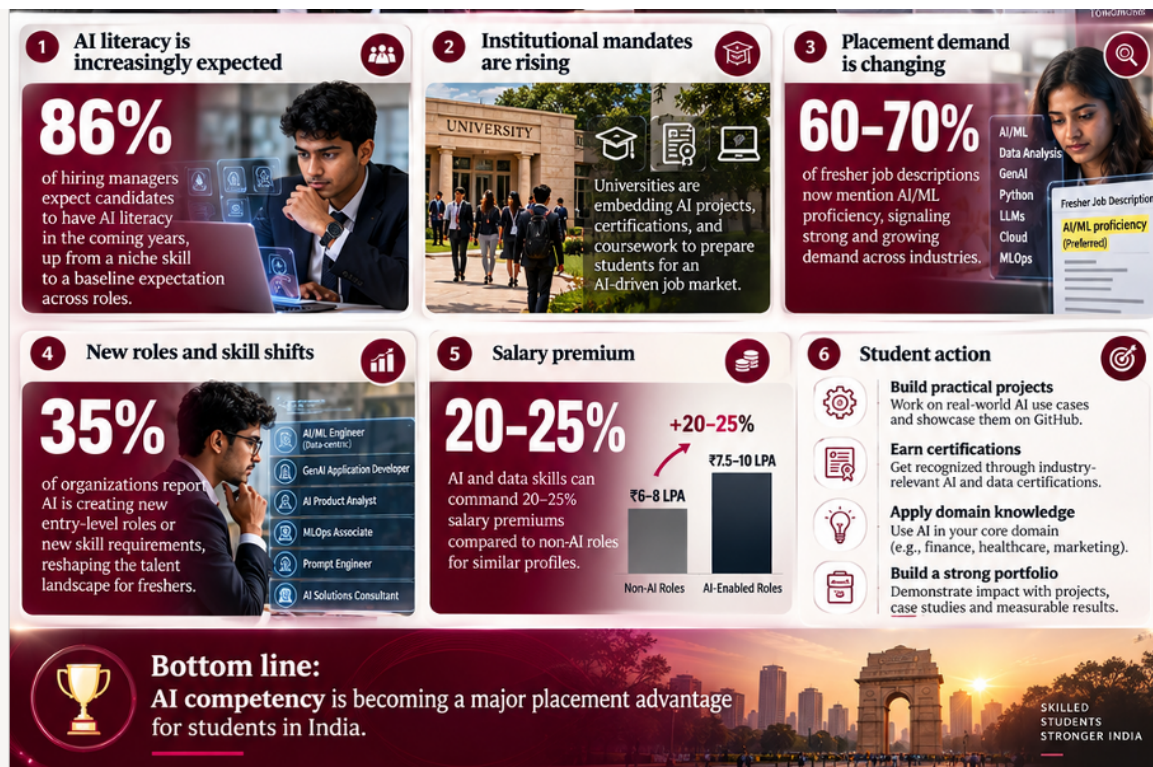
Never read the data right because context limits.
Depending on the 2021

NEW DATA OCT 2026

A8. Claude-related recent-session chart and beginning of the Qwen observation.



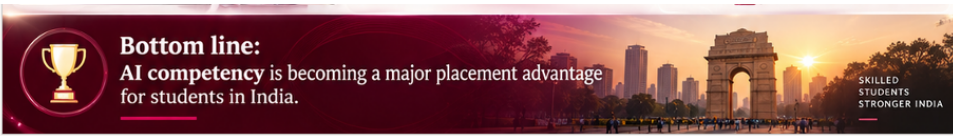
A9. AI Skills & Campus Placements visual; marketing-style claims require primary-source verification.



A10. Placement-claims infographic; percentages are source-provided and unverified in this report.



Bottom line:
AI competency is becoming a major placement advantage for students in India.



SKILLED STUDENTS STRONGER INDIA

Mnemonic 1: A.I.P.L.U.S. (Campus Placements)

AI Skills & Campus Placements: India 2026 — One-Page Report

A	AI literacy is becoming expected	AI skills are increasingly treated as essential for stronger placement outcomes. In many hiring contexts, they are shifting from optional add-ons to expected practical competencies.
I	Institutional mandates are rising	Examples include IILM University's 2026 onward mandatory AI project across programs, AI-for-Managers coursework and Azure AI certification, plus broader curriculum moves such as compulsory AI studies and school-level AI integration from 2026–27.
P	Placement demand is visibly changing	AI/ML proficiency appears in roughly 60–70% of fresher job descriptions, while 35% of organizations say AI is creating new entry-level roles or new skill requirements. Demand is strongest in IT/ITeS, professional services and financial services.
L	Labor-market dynamics are shifting	Traditional mass hiring of generic freshers is declining. Employers increasingly want job-ready candidates who can work with AI tools immediately, while AI also automates boilerplate coding, testing, data cleaning and basic analysis.
U	Upskilling brings measurable premiums	AI and data skills can command 20–25% salary premiums, and analytical/problem-solving skills add about 10–15%. Some campuses are seeing stronger salary bands for students who can demonstrate real AI project work.
S	Student strategy: build proof, not just awareness	Students should focus on practical projects, certifications, domain-specific use cases and ethical AI usage in assessments. Strong candidates can show how they use AI for business, technical or research outcomes—not only that they know the terminology.

Key 2026 Figures:

86% of organizations report AI or agentic AI is transforming campus recruitment; 60–70% of fresher job descriptions mention AI/ML skills; AI/data skills may command 20–25% salary premiums.

A11. Source A.I.P.L.U.S. mnemonic report.

MATERIALS USED

Reference Register

- SINassignment.pdf — supplied 13-page research assignment used as the primary internal source.
- Supplied CHECK framework graphic.
- Supplied M.O.D.E.L.S. framework graphic.
- Supplied blank M.O.D.E.L.S. scoring matrix (TOTAL / 30).
- Supplied SIN School of AI logo and Gayatri Innovation Center image.
- Eleven image artefacts extracted from the supplied assignment PDF.
- Research-team and programme details supplied directly by the requester.

Evidence note

No external web sources were added. Any externally oriented market, employment, pricing or product-version claim should be re-verified against primary sources before public or commercial use.