ChatGPT

# Forecasting Google AI Product Rollouts as of February 4, 2026

## Executive summary

- **Google's near-term AI roadmap is converging on "agentic distribution."** The fastest-moving product surfaces and monetization levers are the agent features being embedded into consumer channels (Search AI Mode, Chrome, Gemini app) and enterprise channels (Vertex AI Agent Builder and Agent Engine), with recurring gating via subscription tier, geography, and "preview" labels. [1]
- **The company's capability roadmap shows a consistent pattern: research disclosure, limited preview, then tier-gated productization.** A clear example is the Genie sequence: an arXiv paper (capability disclosure), DeepMind research previews, then a tier-gated Labs prototype ("Project Genie") for paying subscribers. [2]
- **Case-study calibration: world-model capabilities can go from research preview to consumer prototype in roughly 6 months when positioned as a controlled Labs experience and gated to the highest tier.** Genie 3 was announced August 5, 2025 (limited research preview) and Project Genie began rolling out January 29, 2026 to Google AI Ultra subscribers in the U.S. (18+). [3]
- **"Capability disclosed" versus "SKU shipped" is a core mismatch to track.** Genie 3 disclosed "promptable world events," but Project Genie explicitly notes those events are "not yet included," and the prototype is capped at 60 seconds. This pattern is also visible across Gemini 3: models are broadly surfaced, but key elements remain in preview with explicit constraints and tier limits. [4]
- **Most likely rollouts over the next 6 months (to August 4, 2026):**
- Expansion of Project Genie beyond the U.S. and iterative addition of withheld Genie 3 features (notably "promptable events") inside the prototype experience. [5]
- Continued geographic and tier expansion of Gemini Agent and Chrome "auto browse," paired with deeper app-connectivity via Personal Intelligence. [6]
- More "preview-to-GA" transitions in enterprise agent infrastructure (sessions, memory, code execution) driven by billing start dates and region rollout patterns. [7]
- **Most important structural risk for the roadmap:** high-cost modalities (video, world simulation, full-browser agents) remain compute- and safety-constrained, which increases the probability that rollouts stay behind paywalls, remain region-limited, or ship as "preview/experimental" longer than investors may expect. Evidence of this constraint shows up repeatedly via quotas, tier-specific limits, and prototype caps. [8]

## Capability cluster taxonomy and leading indicators

The taxonomy below is organized as "capability clusters" with: what the capability is, pre-product research artifacts, where it lands, high-signal leading indicators, and empirically grounded lag ranges based on recent Google examples.

| Cluster | Capability definition | Pre-product artifacts that typically precede productization | Primary product surface area | Leading indicators and “tells” (high signal first) | Typical lag ranges with empirical anchors |
|---|---|---|---|---|---|
| Foundation LLM and multimodal reasoning | Frontier general models that reason across text, images, audio, video and code; includes “thinking” controls, long context, and tool use. [9] | Model-family blog posts, benchmark disclosures, developer guides, model cards, API model IDs and deprecation tables. [10] | Consumer: Gemini app and Search AI Mode. Enterprise and dev: Vertex AI and Gemini API/ AI Studio. [11] | Official “preview/stable” status shifts and model IDs; API changelog entries for launches/ deprecations; lifecycle pages with discontinuation dates; pricing/ billing start for features like grounding. [12] | Experimental or preview to broad availability can compress to months when the API surface is ready (Gemini 2.5 disclosed Mar 25, 2025 and moved into broader rollout and pricing cycles by June 2025). [13] |
| Generative media (image, video, audio) | Media generation, editing and post-production controls (e.g., image generation, video generation, native audio). Provenance controls (watermarking) are tightly coupled. [14] | DeepMind model pages, public previews, API docs for generation endpoints, product demos and later GA announcements. [15] | Consumer: Gemini app, Labs tools (Flow, Whisk). Enterprise and dev: Vertex AI, Gemini API. [16] | Model-family announcements tied to I/O; availability in AI Studio and Vertex; “paid preview” language; new parameters (aspect ratio, upsampling); SynthID watermark mentions. [17] | Announce to API GA can be ~3 months (Imagen 4 introduced May 20, 2025, GA in Gemini API Aug 15, 2025). [18] |

| Cluster | Capability definition | Pre-product artifacts that typically precede productization | Primary product surface area | Leading indicators and “tells” (high signal first) | Typical lag ranges with empirical anchors |
|---|---|---|---|---|---|
| World models and interactive simulation | Models that generate interactive, action-controllable environments with temporal consistency; often positioned for agent training and creative tools. [19] | ArXiv papers (world models), DeepMind research posts, embodied agent reports demonstrating use, limited research previews. [20] | Consumer: Labs prototypes like Project Genie. Research and future enterprise: agent training and simulation proof points via embodied agents. [21] | “Limited research preview” language; explicit controls and limitations; a Labs CTA appearing on research posts; prototype caps (duration, action space) and tier gating. [22] | Research preview to consumer prototype can be ~6 months when tightly gated (Genie 3 announced Aug 5, 2025; Project Genie rollout Jan 29, 2026). [23] |
| Agents (tool use, long-horizon tasks, autonomy constraints) | Systems that plan, call tools, browse, and take multi-step actions under supervision and safety constraints. Includes scheduling and connected-app actions. [24] | Agent demos, safety policies, developer docs for tool calling and MCP, enterprise agent platform release notes (sessions, memory, code execution). [25] | Consumer: Gemini Agent in Gemini app; auto browse in Chrome; scheduled actions in Gemini apps; Search AI Mode interactive tools. Enterprise: Agent Builder/ Engine. [26] | Clear gating to Ultra/Pro, U.S.-only launches, age requirements; “experimental/ preview” labels; billing start dates for agent primitives; replacement of older tool interfaces with agent mode + MCP. [27] | Enterprise agent primitives show staged rollout with monetization gates (sessions/ memory GA and charging dates in late 2025 to early 2026). [28] |

| Cluster | Capability definition | Pre-product artifacts that typically precede productization | Primary product surface area | Leading indicators and “tells” (high signal first) | Typical lag ranges with empirical anchors |
|---|---|---|---|---|---|
| Embodied AI and robotics | Vision-language-action and embodied reasoning models that plan and act in physical contexts; often staged through trusted testers and developer previews before broad use. [29] | DeepMind robotics posts and technical reports, partner announcements, API previews for robotics reasoning models. [30] | Research and dev: Robotics-ER in Gemini API/AI Studio; longer-term enterprise and partner deployments. [31] | “Selected trusted testers” wording; previews in Gemini API; tech report publication; explicit safety-controller interfacing language. [32] | Initial foundation disclosure to dev preview can be ~6 months (Gemini Robotics announced Mar 12, 2025; Robotics-ER 1.5 preview appears in Gemini API documentation and changelogs later in 2025). [33] |
| Serving and cost curves (inference optimization, latency, edge) | Pricing, caching, batch processing, “thinking budget,” and model lifecycle management. This is a critical gate for broad rollouts. [34] | Pricing pages, model lifecycle docs, batch APIs, caching and quota docs, deprecation notices. [35] | Dev and enterprise first (Gemini API and Vertex), then consumer limits/credits and tier quotas. [36] | Pricing deltas and new cost levers (context caching, batch); explicit discontinuation dates; deprecation tables; new “thinking level” knobs exposed in APIs. [37] | Preview to stable often correlates with mature cost controls (batch, caching) and explicit lifecycle commitments. [38] |

| Cluster | Capability definition | Pre-product artifacts that typically precede productization | Primary product surface area | Leading indicators and “tells” (high signal first) | Typical lag ranges with empirical anchors |
|---|---|---|---|---|---|
| Developer platform and monetization (APIs, pricing, tiering, gating) | Unification of SDKs, API surface expansion, tier-based entitlements, bundling with cloud credits, and billing start for features. [39] | Official SDK repositories, migration guides, API changelogs, “preview vs GA” lifecycle docs, cloud release notes. [40] | Developers: Gemini API and AI Studio; enterprise: Vertex AI; subscription entitlements connecting consumer plans to dev benefits. [41] | SDK unification and deprecations (legacy SDKs archived); model string changes; “billing begins on X date” changelog notes; cloud credits bundled into subscriptions. [42] | API feature launch to billing start can be weeks to months (e.g., grounding billing start dates in Gemini API changelog). [43] |
| Consumer distribution channels (apps, subscriptions, Workspace, Android) | Shipping AI capabilities at scale through Search, Chrome, Gemini app, Android ecosystem devices, and subscription plans. [44] | Product posts, support docs, rollout notes, subscription pages and country availability lists. [45] | Search AI Mode, Chrome, Gemini app, Android Auto/Wear/TV, home devices. [46] | Geographic rollout language (“U.S. first,” “expanding in due course”), account-type restrictions, tiered quotas/credits, and explicit “coming soon” product-page status. [47] | |

| Cluster | Capability definition | Pre-product artifacts that typically precede productization | Primary product surface area | Leading indicators and “tells” (high signal first) | Typical lag ranges with empirical anchors |
|---|---|---|---|---|---|
| Safety, provenance, and policy controls | Watermarking and provenance (SynthID), tool-use safety constraints, and policy language that limits or shapes rollout. [48] | Responsibility sections in research/blog posts, watermark adoption in docs, support docs describing user confirmations, and enterprise governance controls. [49] | Cross-cutting across consumer and enterprise, often as gating. [50] | Explicit limitations, confirmation prompts, watermark requirements, and staged access (“trusted testers,” “preview,” “Labs”). [51] | Safety review lengthens lags for high-risk modalities (agents, world models), frequently resulting in tier, region, and feature caps. [52] |

# Research-to-product conveyor model

## Conveyor stages and gating mechanisms

This conveyor model reflects the most repeatedly observed shipping pattern across Google AI surfaces from roughly 2024 to early 2026, emphasizing how “capability disclosed” becomes a “productized SKU shipped.”

- **Stage: Research disclosure (capability disclosed)**
  Typical artifacts include arXiv papers (e.g., Genie) and DeepMind research posts (e.g., Genie 3). [53]
  Gating at this stage is primarily narrative and safety framing (“limitations,” “responsibility,” and “research preview” language). [54]

- **Stage: Limited research preview (capability accessible to a small cohort)**
  This stage is explicitly described for Genie 3 as a “limited research preview” for a small cohort of academics and creators. [54]
  Leading indicators include the presence of a “Try” CTA in research posts and references to expanding to additional testers in the future. [55]

- **Stage: Prototype productization in a controlled surface (SKU shipped, but constrained)**
  For world models, this appears as a Google Labs prototype gated to a top subscription tier, with explicit caps and withheld features. Project Genie is explicitly positioned this way, including a 60-second limit and missing “promptable events.” [56]

- **Stage: Developer preview (SKU-like access through APIs, still guarded)**
  This shows up via “preview” model IDs in the Gemini API Developer Guide and “public preview” in Vertex AI release notes. [57]

The Gemini API explicitly categorizes models as stable, preview, latest, or experimental, and model string naming conventions become a key tell. [58]

- **Stage: Enterprise readiness (guardrails + reliability + governance)**
  Vertex AI adds lifecycle metadata (release and discontinuation dates), security controls, and structured deprecation. [59]
  Agent platform primitives also move through monetization gates such as "begin charging for usage on" specific dates. [28]

- **Stage: Broad consumer rollout (scale distribution)**
  Search and Chrome show a pattern of U.S.-first + subscribers-first rollouts followed by stated intent to expand to "everyone" or more regions. For example, Search AI Mode with Gemini 3 launched to Pro and Ultra subscribers in the U.S. with plans to bring Gemini 3 in AI Mode to everyone in the U.S. [60]

## Empirical disclosure-to-rollout dataset

The cases below provide a light-weight "dataset" of disclosure-to-rollout pathways with observable lags and gating.

| Disclosure or key milestone | Productized surface milestone | What it implies about lags and gating |
|---|---|---|
| Genie paper submitted Feb 23, 2024 (capability disclosed). [61] | Project Genie rollout began Jan 29, 2026 for top-tier subscribers in the U.S. (SKU shipped, constrained). [62] | "Paper to consumer prototype" can be roughly 1.5 to 2 years for frontier world models; early consumer shipping uses strict caps and tier gating. [63] |
| Genie 3 announced Aug 5, 2025 with real-time interaction at 24 fps, 720p, and "few minutes" consistency. [64] | Project Genie prototype: 60-second generation limit and missing promptable events. [62] | Research preview to constrained Labs prototype can be about 6 months; withheld features are signaled explicitly. [23] |
| Imagen 4 introduced May 20, 2025. [65] | Imagen 4 GA in Gemini API and AI Studio Aug 15, 2025. [66] | Media models can transition from announcement to API GA in ~3 months when safety and cost curves are manageable and endpoints stabilize. [18] |
| Gemini 2.5 introduced Mar 25, 2025 as "thinking models," with availability in AI Studio and app. [67] | Gemini 2.5 Pro GA in Vertex reported with a June 17, 2025 release date and a June 17, 2026 discontinuation date. [68] | Enterprise lifecycle controls become explicit once models are productized; discontinuation dates can serve as leading indicators for refresh cycles. [69] |

| Disclosure or key milestone | Productized surface milestone | What it implies about lags and gating |
|---|---|---|
| Gemini Robotics and Robotics-ER introduced Mar 12, 2025. [70] | Robotics-ER appears as a preview model in Gemini API documentation. [71] | Embodied AI commonly routes through "trusted testers" and preview gating before broader access. [32] |
| Gemini Code Assist "tools" deprecated Oct 14, 2025 and replaced by agent mode using MCP servers. [72] | Vertex AI Agent Builder and Agent Engine continue staged GA and billing transitions through late 2025 and early 2026. [28] | Google is converging on MCP-like tool ecosystems across products; deprecations are a key signal of "platform direction." [73] |

### Core inference about the pipeline

Across clusters, the highest predictive signals are the **mechanisms that require operational commitments**: model lifecycle dates, billing start dates, deprecation tables, and subscription entitlement pages. These are harder to "hand-wave" than demos or press quotes and thus tend to be higher signal for near-term rollouts. [74]

## Case study: Project Genie and Genie 3

### Timeline reconstruction

The sequence below is the required calibration chain, built from primary sources plus reputable press.

- **February 23, 2024: Genie paper (capability disclosed).**
  The arXiv abstract describes Genie as an "unsupervised" generative interactive environment trained from unlabeled internet video, capable of generating "action-controllable virtual worlds" and framed as a "foundation world model." [61]

- **December 4, 2024: Genie 2 research post (capability step-up).**
  DeepMind positioned Genie 2 as a "large-scale foundation world model," an intermediary step before Genie 3's real-time interactivity. [75]

- **August 5, 2025: Genie 3 research preview (capability disclosed plus access gating).**
  DeepMind states Genie 3 can be navigated "in real time at 24 frames per second," retains consistency "for a few minutes," and operates at "720p." [64]
  The same post explicitly frames Genie 3 as a "limited research preview," giving early access to "a small cohort of academics and creators." [54]

- **December 2025: SIMA 2 technical report links agent training and world models, explicitly referencing Genie 3.**
  The SIMA 2 report states the agent generalizes to "photorealistic worlds generated on-the-fly by Genie 3," and describes Genie 3 as a generative world model enabling real-time interaction in newly-created environments. [76]

- **January 29, 2026: Project Genie launches as a consumer-accessible prototype (SKU shipped, constrained).**
  The Project Genie post positions it as a "prototype web app" powered by Genie 3 plus image generation and Gemini, and begins rollout to Google AI Ultra subscribers in the U.S. (18+), with expansion "to more territories in due course." [62]
  Packaging is explicit: it is "an experimental research prototype in Google Labs," with a 60-second generation limit and missing Genie 3 capabilities such as promptable events. [62]

- **January 29 to 30, 2026: Reputable press confirms distribution posture and market impact.**
  TechCrunch frames Project Genie as a top-tier Labs prototype and contextualizes competitive dynamics.
  Reuters notes the Project Genie prototype access and reports market reactions (video game stocks declining on the announcement). [77]

## Leading indicators observed before broader distribution

The "tells" that appeared before Project Genie became available to paying users are unusually clean in this chain:

- **Explicit access language and cohort definition in the research post** ("limited research preview" plus "small cohort"). This is a high-signal leading indicator because it implies operational readiness for controlled distribution and feedback collection, not just a paper release. [54]
- **A direct "Try Project Genie" call-to-action from within the Genie 3 research post** is an unusually strong bridging artifact between research and product surface. [64]
- **Agent-walkthrough adjacency:** Genie 3's post emphasizes that it was used to generate worlds for a version of the SIMA agent, and the SIMA 2 report reinforces this coupling. That pairing suggests "world model + agent" is a key internal product thesis. [78]

## Packaging decisions and gating

Project Genie's packaging reflects a repeatable approach for high-cost, high-risk capabilities:

- **Tier gating:** access begins with top-tier subscribers in a single geography. [62]
- **Geographic and demographic gating:** U.S.-only and 18+ requirement. [62]
- **Account-type exclusions:** explicitly "not available" to "Ultra for Business" users. [62]
- **Scope-limited feature exposure:** an explicit 60-second generation cap and disclosure that promptable events are not yet included. [62]
- **Toolchain coupling:** prototype is powered by Genie 3, image generation and the Gemini assistant, reflecting "capability bundling" as a productization strategy rather than shipping a raw Genie API first. [79]

## Calibration takeaways for forecasting

- **Lag distribution estimate for world models:** for expensive simulation-like capabilities, a reasonable base-case lag from research preview to constrained consumer prototype is ~4 to 9 months, anchored by the Genie 3 to Project Genie transition. [23]
- **Signal reliability ranking:** "limited research preview" language plus a product CTA is materially higher signal than press chatter or benchmark-only claims. [54]

- **Business-model inference:** top-tier subscribers are being used as a monetized test cohort for frontier capabilities that are both expensive (compute) and sensitive (safety and misuse). [80]

## Roadmap forecast for the next 6, 12, and 18 months

The roadmap below uses probability bands (60 to 80 percent, 40 to 60 percent, 20 to 40 percent) and explicit trigger conditions. Each item is labeled as either **capability disclosed** (research or preview exists) or **SKU shipped** (a product surface exists, but may be constrained).

### Ranked forecast items with probabilities and triggers

| Forecasted rollout | Cluster | Capability disclosed vs SKU shipped | Probability | Expected window | Key triggers that raise confidence | Key blockers and failure modes |
|---|---|---|---|---|---|---|
| Expansion of Project Genie access beyond the U.S., and possibly beyond Ultra-only gating | World models; consumer channels | SKU shipped (constrained); expansion pending | 60 to 80 percent | 6 to 12 months | Country availability list or subscription entitlement updates; explicit “expanding to more territories” execution; Labs rollout notes. [81] | Compute cost, safety concerns, IP and likeness risks, misuse via world simulation; could remain Ultra-only longer. [5] |
| Project Genie feature upgrades: promptable events and longer interaction horizons | World models | Capability disclosed; partial SKU shipped | 40 to 60 percent | 6 to 12 months | Product update notes or DeepMind “future updates” referenced in Project Genie; removal or raise of the 60-second cap; explicit mention that promptable events are now included. [56] | Reliability, latency, multi-agent safety, real-world-location accuracy limits; could ship slowly with small expansions. [82] |

| Forecasted rollout | Cluster | Capability disclosed vs SKU shipped | Probability | Expected window | Key triggers that raise confidence | Key blockers and failure modes |
|---|---|---|---|---|---|---|
| A developer-accessible Genie or world-model API in a limited preview form (likely via Vertex or a gated program rather than fully open) | World models; developer platform | Capability disclosed; SKU not yet shipped in API form | 20 to 40 percent | 12 to 18 months | New model IDs or endpoints in Vertex model lists; “preview” world model endpoints; SDK commits referencing world simulation. [83] | Safety, misuse, and cost; unclear enterprise demand clarity; may stay as internal research substrate for agents rather than external API. [84] |
| Wider rollout of Gemini Agent beyond Ultra U.S. English, and integration into additional Google products | Agents; consumer channels | SKU shipped (limited) | 60 to 80 percent | 6 to 12 months | Gemini Agent country/ language FAQ changes; support docs expanding eligibility; new connected apps list; more regions in product posts. [85] | Agent error risk, prompt injection, transactional safety, regulatory scrutiny for automated actions; could remain supervised and region-limited. [86] |
| Chrome “auto browse” expansion from U.S. Pro/Ultra to more regions and tighter Personal Intelligence integration | Agents; consumer channels | SKU shipped (U.S. first) | 60 to 80 percent | 6 to 12 months | Chrome post indicates “coming months” type expansion; Personal Intelligence explicitly positioned as cross-app context; more locales and tiers listed. [87] | Security and adversarial web risks; user trust issues; gating may remain to reduce liability and cost. [88] |

| Forecasted rollout | Cluster | Capability disclosed vs SKU shipped | Probability | Expected window | Key triggers that raise confidence | Key blockers and failure modes |
|---|---|---|---|---|---|---|
| Search AI Mode scaling: Gemini 3 routing for harder queries expands, with generative UI tools and simulations becoming more common | Consumer channels; foundation LLM | SKU shipped (U.S. subscribers first) | 60 to 80 percent | 6 to 12 months | Search post describes “coming weeks” for routing and “soon” for broader U.S. rollout; watch for expanded eligibility lists and UI feature standardization. [89] | Regulatory and publisher issues, quality regressions, hallucination risk; could slow expansion or restrict UI tools. [90] |
| Gemini 3 family transitions from “preview” to more stable/GA positioning in developer and enterprise surfaces, paired with retirement of older model generations | Serving/ cost; developer platform | Capability shipped as preview; GA not yet complete | 40 to 60 percent | 6 to 12 months | “All models currently in preview” changes; Vertex “launch stage: GA” updates; acceleration near discontinuation dates for 2.5 models; deprecation notices. [91] | Cost and safety gating, compliance needs; could extend preview if reliability not stable for enterprises. [92] |

| Forecasted rollout | Cluster | Capability disclosed vs SKU shipped | Probability | Expected window | Key triggers that raise confidence | Key blockers and failure modes |
|---|---|---|---|---|---|---|
| Vertex AI Agent Builder and Agent Engine platform hardening: more GA transitions, broader region coverage, and charging milestones | Agents; developer platform | SKU shipped; platform maturing | 60 to 80 percent | 6 to 12 months | Release note cadence: sessions/ memory GA, explicit “begin charging” dates, “preview to GA” motions for design and registry. [28] | Enterprise adoption friction, cost objections, competitive pressure; could lead to pricing changes or slower GA progression. [93] |
| Veo 3.1: continued production hardening (controls, aspect ratios, upsampling) and broader developer availability as pricing and tooling mature | Generative media; serving/ cost | SKU shipped (paid preview) | 60 to 80 percent | 6 to 12 months | Vertex release notes already show capability updates; watch for GA language and broader quotas; Gemini API posts and demo apps act as leading indicators. [94] | Copyright and misuse risks, high GPU costs; may remain paid preview longer than text models. [95] |

| Forecasted rollout | Cluster | Capability disclosed vs SKU shipped | Probability | Expected window | Key triggers that raise confidence | Key blockers and failure modes |
|---|---|---|---|---|---|---|
| Robotics-ER and embodied reasoning models migrate from preview toward broader developer access, with stronger “thinking budget” controls | Embodied AI and robotics; serving/ cost | Capability shipped as preview | 40 to 60 percent | 12 to 18 months | Preview-to-GA signals in Gemini API docs; expansion of robotics docs and samples; partner announcements and toolchains. [96] | Physical safety and product liability; limited immediate consumer distribution; may remain research and partner-led. [32] |
| Subscription packaging evolution: AI Plus expansion suggests further tier reshaping and bundling, potentially bringing more “preview” capabilities to mid-tier plans with strict quotas | Developer platform; consumer channels | SKU shipped; packaging evolving | 40 to 60 percent | 6 to 12 months | Country availability expansions and benefit bundling (AI Plus rollout, cloud credits in Pro/Ultra); watch for entitlement table changes and quota adjustments. [97] | Compute economics could push the opposite direction: tightening quotas, raising prices, or keeping frontier features Ultra-only. [98] |

## Scenario framing

- **Bearish scenario: rollouts slow and consolidate behind paywalls**
  Probability increases if: billing milestones accelerate for high-cost agent primitives and media generation, and consumer quotas tighten to protect margins. Signals consistent with this include explicit usage limits, preview caps, and “begin charging” dates across enterprise agent tooling. [99]

- **Competitive acceleration scenario: rollouts expand faster to defend distribution moats**
  Probability increases if: Search AI Mode and Chrome agents broaden rapidly beyond paid tiers, and Personal Intelligence expands to become the default mode of interaction across app surfaces. Early evidence for distribution-driven acceleration includes Search launching Gemini 3 “day one” in AI Mode and explicit “soon” language for wider U.S. rollout. [100]

# Monitoring plan and signal stack

## Signal stack tiers with weights and validation rules

- **Tier A (weight: highest): official primary sources**
  Examples: product posts and official documentation for Search AI Mode, Chrome auto browse, Gemini app/Agent, subscription entitlements, and Cloud release notes. [101]
  Actionable when: wording changes from “preview” to “GA,” a billing start date is added, or country and account eligibility expands. [102]
  Validation: treat the latest official doc as source of truth, and reconcile conflicts by timestamp. [103]

- **Tier B (weight: high): technical artifacts**
  Examples: arXiv papers and technical reports (Genie, SIMA 2), model cards, and DeepMind research posts specifying limitations and evaluation framing. [104]
  Actionable when: a technical report introduces a new environment class, interface, or capability that connects to an existing product surface. SIMA 2’s explicit linkage between Genie 3 environments and generalist agents is a strong “two-cluster coupling” indicator. [105]
  Validation: confirm the artifact is authored by the relevant lab and that subsequent Tier A docs reflect packaging or constraints consistent with the capability. [56]

- **Tier C (weight: medium-high): developer surface changes**
  Examples: Gemini API changelog and model IDs, Vertex AI release notes, SDK repository migrations and release notes. [106]
  Actionable when: a model ID appears in “preview” before marketing posts, billing starts for a feature, a legacy SDK is deprecated, or a new endpoint replaces a preview endpoint. [107]
  Validation: require at least one Tier A confirmation before upgrading confidence on a “rollout” claim, especially for features found only via code or changelogs. [108]

- **Tier D (weight: low): community chatter and press leads**
  Examples: Hacker News threads for Genie 3 and Project Genie.
  Actionable when: a credible leak points to a model ID, a hidden entitlement, or a region rollout before official posts.
  Validation: never treat as confirmation; use only to generate hypotheses, then confirm via Tier A or Tier C. [109]

## Watchlist: sources, keywords, repositories, and author groups

High-signal sources to monitor weekly:

- “Models and Research” and product posts for expansions of Search AI Mode, Chrome, Gemini app, and Labs experiments. [110]

- Gemini API changelog and model list pages for new model IDs, deprecations, and billing start notices. [111]
- Vertex AI release notes for preview-to-GA transitions, regional expansions, and charging milestones for agent infrastructure. [28]
- Official SDK repos and releases, especially the unified Gen AI SDK and migration signals from deprecated libraries. [112]

Keywords and patterns (high signal):

- “begin charging,” “billing starts,” “deprecated,” “shut down,” “discontinuation date,” “launch stage: GA,” “public preview,” “limited research preview,” “rolling out,” “expanding to more territories,” and new model IDs that include “preview” or new date-stamped suffixes. [113]

Author-group cues (useful, but secondary to Tier A/C signals):

- World models and embodied agent intersection: the Genie and SIMA teams are explicitly linked via technical acknowledgments and cross-references, so new papers or posts in either team can foreshadow product evolutions in world simulation and agent training. [114]

### Roadmap radar checklist

A weekly cadence that a human analyst can run:

- Check official entitlement and rollout changes: subscription plan pages and Help Center quota tables. Red flag: sudden quota tightening or removal of features. High-confidence indicator: new regions listed for formerly U.S.-only features. [115]
- Check gemini.google release notes for newly shipped “what” items and dates. High-confidence indicator: an entry stating “Starting today” for a previously “coming soon” feature. [116]
- Check Gemini API changelog for new model IDs and billing start dates. High-confidence indicator: a model ID moves from preview naming to stable naming, or billing begins for a feature that was previously free. [111]
- Check Vertex AI release notes for agent platform milestones and cost changes. High-confidence indicator: preview to GA transitions, new regions, and adjusted charging dates. [28]
- Scan Tier D only after the above: Hacker News and similar sources to detect early rumors. Action only if a Tier C artifact appears within 1 to 2 weeks. [117]

## Evidence appendix

The table below compiles key claims used in the methodology and the forecasts, with supporting sources, cluster mapping, indicator tier, and confidence.

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Genie paper submitted Feb 23, 2024 describes Genie as an unsupervised “generative interactive environment” trained from unlabeled internet video and capable of “action-controllable virtual worlds.” | 61 | World models | B | High | Foundational capability disclosure; not the same as product access. |
| Genie 3 announced Aug 5, 2025 can be navigated in real time at 24 fps, retains consistency for a few minutes, and operates at 720p. | 64 | World models | A | High | Performance characteristics are declared in the research post; real-world product constraints later differ. |
| Genie 3 is explicitly a “limited research preview” for a small cohort of academics and creators. | 54 | World models | A | High | Strong leading indicator of impending controlled product surface. |
| SIMA 2 technical report states the agent generalizes to “photorealistic worlds generated on-the-fly by Genie 3” and describes Genie 3 as enabling real-time interaction in newly created environments. | 76 | Agents; world models | B | High | Confirms cross-cluster coupling: agent architecture plus world model generation. |

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Project Genie began rolling out Jan 29, 2026 to Google AI Ultra subscribers in the U.S. (18+), expanding to more territories later. | 62 | World models; consumer channels | A | High | “Rollout begins” is explicit; expansion timing is not guaranteed. |
| Project Genie is positioned as an experimental prototype in Google Labs and uses Genie 3 plus image generation and Gemini. | 62 | World models; developer platform | A | High | Important sign of productization approach: bundling capabilities rather than exposing raw world-model APIs. |
| Project Genie has a 60-second generation limit and does not yet include Genie 3 “promptable events.” | 62 | World models | A | High | “Capability disclosed vs SKU shipped” mismatch is explicit. |
| Reputable press corroborates Project Genie’s prototype positioning and discusses market impact (gaming stocks reaction). | 118 | World models | D (press) | Medium | Use for context and market reaction; primary product facts still come from Google posts. |
| Gemini API docs define stable, preview, latest, and experimental model versions and describe model string naming patterns. | 58 | Developer platform | A | High | Key for monitoring: appearance of model IDs commonly precedes broad marketing. |

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Gemini 3 Developer Guide states all Gemini 3 models are currently in preview and lists model IDs and pricing. | 119 | Foundation LLM; developer platform | A | High | Strong signal that “GA” transitions remain ahead for some surfaces. |
| Gemini API changelog documents Gemini 3 Pro preview launch and subsequent deprecations and billing notices (e.g., grounding billing begins Jan 5, 2026). | 43 | Developer platform; serving/cost | C | High | Billing-start notes are high-signal monetization and maturity indicators. |
| Search AI Mode: Gemini 3 shipped “day one” in Search AI Mode; Pro and Ultra U.S. subscribers can select Gemini 3 Pro, and Google states it will bring Gemini 3 in AI Mode to everyone in the U.S. “soon.” | 89 | Consumer channels | A | High | This is a direct Tier A rollout statement; “soon” is directional, not a promise. |
| Chrome auto browse is introduced for Pro and Ultra U.S. subscribers, and Google signals “Personal Intelligence” is coming to Chrome in the coming months. | 120 | Agents; consumer channels | A | High | Exact geography and tier expansion beyond U.S. not guaranteed. |

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Gemini Agent: available with Gemini 3 Pro; rolling out to Google AI Ultra subscribers in the U.S. with language set to English; Workspace and student accounts excluded for now. | 121 | Agents | A | High | Strong base for forecasting expansion once reliability and safety mature. |
| Support docs describe scheduled actions entitlements, requiring Pro or Ultra for personal accounts, and cap at 10 active actions. | 122 | Agents | A | High | Indicates constraints that may persist until reliability improves. |
| Gemini Apps quota tables show tier-based limits for images and videos and continued “preview” labeling for Veo 3.1 features. | 123 | Serving/cost; consumer channels | A | High | High-signal for compute constraints and pricing pressure. |
| Build-with-Gemini-3-Flash post states context caching and Batch API cost reductions and lists pricing, indicating maturity steps for cost curves. | 124 | Serving/cost | A | High | Cost levers often precede broader rollouts. |
| Vertex release notes show Agent Builder and Agent Engine milestones, including GA transitions for sessions/memory and charging date adjustments. | 28 | Agents; developer platform | C | High | Billing and GA transitions are among the most reliable rollout predictors. |

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Imagen 4 introduced May 20, 2025 and later GA in Gemini API and AI Studio Aug 15, 2025. | 18 | Generative media | A | High | Empirical anchor for “announcement to GA” compression. |
| Veo 3.1 and 3.1 Fast are released in paid preview in the Gemini API and also available via AI Studio and Vertex, with ongoing capability updates. | 125 | Generative media | A/C | High | Paid preview language is a precursor to GA, not a guarantee of near-term broad access. |
| Forecast: Project Genie expands to more territories in the next 6 to 12 months. | 126 | World models | Inference from A | Medium | Evidence: explicit “expanding to more territories” plus observed precedent of expanding AI Plus availability. Timing window remains probabilistic. |
| Forecast: Project Genie adds promptable events and extends the 60-second cap within 6 to 12 months. | 56 | World models | Inference from A | Medium | Evidence: Genie 3 capability disclosed; Project Genie notes it is “not yet included.” Could slip if safety or latency issues persist. |
| Forecast: Gemini Agent expands beyond Ultra U.S. English within 6 to 12 months. | 127 | Agents | Inference from A | Medium | Evidence: repeated “expand soon” language and pattern of U.S.-first scaling in Search AI Mode and Assistant transition posts; still constrained by safety and reliability. |

| Claim | Source(s) | Cluster | Indicator type (A/B/C/D) | Confidence | Notes and caveats |
|---|---|---|---|---|---|
| Forecast: a developer-accessible world-model API appears in limited preview within 12 to 18 months. | 128 | World models | Inference from A/B/C | Low to Medium | Evidence is indirect: coupling of agents + world models and platform direction toward tool ecosystems. Major blocker is misuse and cost. |
| Hacker News threads are useful noisy leads for early reaction and possible sightings, but require confirmation via Tier A or Tier C. | | Monitoring | D | Medium | Treat purely as hypothesis generation given high false-positive rates. |


1 11 44 46 60 89 90 100 101 110 https://blog.google/products-and-platforms/products/search/gemini-3-search-ai-mode/
https://blog.google/products-and-platforms/products/search/gemini-3-search-ai-mode/

2 19 20 53 61 63 104 https://arxiv.org/abs/2402.15391
https://arxiv.org/abs/2402.15391

3 4 22 23 49 51 52 54 55 64 78 82 84 https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/
https://deepmind.google/blog/genie-3-a-new-frontier-for-world-models/

5 21 47 50 56 62 79 80 81 126 https://blog.google/innovation-and-ai/models-and-research/google-deepmind/project-genie/
https://blog.google/innovation-and-ai/models-and-research/google-deepmind/project-genie/

6 24 26 27 85 86 88 121 127 https://gemini.google/overview/agent/
https://gemini.google/overview/agent/

7 28 93 94 102 103 113 https://docs.cloud.google.com/vertex-ai/docs/release-notes
https://docs.cloud.google.com/vertex-ai/docs/release-notes

8 95 98 99 115 123 https://support.google.com/gemini/answer/16275805?hl=en
https://support.google.com/gemini/answer/16275805?hl=en

9 10 13 67 69 https://blog.google/innovation-and-ai/models-and-research/google-deepmind/gemini-model-thinking-updates-march-2025/
https://blog.google/innovation-and-ai/models-and-research/google-deepmind/gemini-model-thinking-updates-march-2025/

12 58 83 109 https://ai.google.dev/gemini-api/docs/models
https://ai.google.dev/gemini-api/docs/models

14 15 66 https://developers.googleblog.com/announcing-imagen-4-fast-and-imagen-4-family-generally-available-in-the-gemini-api/
https://developers.googleblog.com/announcing-imagen-4-fast-and-imagen-4-family-generally-available-in-the-gemini-api/

16 17 18 65 https://blog.google/innovation-and-ai/products/generative-media-models-io-2025/
https://blog.google/innovation-and-ai/products/generative-media-models-io-2025/

25 72 73 https://docs.cloud.google.com/gemini/docs/release-notes
https://docs.cloud.google.com/gemini/docs/release-notes

29 30 32 33 70 https://deepmind.google/blog/gemini-robotics-brings-ai-into-the-physical-world/
https://deepmind.google/blog/gemini-robotics-brings-ai-into-the-physical-world/

31 71 96 https://ai.google.dev/gemini-api/docs/robotics-overview
https://ai.google.dev/gemini-api/docs/robotics-overview

34 35 36 37 38 124 https://blog.google/innovation-and-ai/technology/developers-tools/build-with-gemini-3-flash/
https://blog.google/innovation-and-ai/technology/developers-tools/build-with-gemini-3-flash/

39 https://ai.google.dev/gemini-api/docs/libraries
https://ai.google.dev/gemini-api/docs/libraries

40 112 https://github.com/googleapis/python-genai
https://github.com/googleapis/python-genai

41 https://docs.cloud.google.com/vertex-ai/generative-ai/docs/sdks/overview
https://docs.cloud.google.com/vertex-ai/generative-ai/docs/sdks/overview

42 https://github.com/google-gemini/deprecated-generative-ai-js
https://github.com/google-gemini/deprecated-generative-ai-js

43 106 107 108 111 117 128 https://ai.google.dev/gemini-api/docs/changelog
https://ai.google.dev/gemini-api/docs/changelog

45 97 https://blog.google/products-and-platforms/products/google-one/google-ai-plus-availability/
https://blog.google/products-and-platforms/products/google-one/google-ai-plus-availability/

48 https://ai.google.dev/gemini-api/docs/image-generation
https://ai.google.dev/gemini-api/docs/image-generation

57 91 119 https://ai.google.dev/gemini-api/docs/gemini-3
https://ai.google.dev/gemini-api/docs/gemini-3

59 68 74 https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-pro
https://docs.cloud.google.com/vertex-ai/generative-ai/docs/models/gemini/2-5-pro

75 https://deepmind.google/blog/genie-2-a-large-scale-foundation-world-model/
https://deepmind.google/blog/genie-2-a-large-scale-foundation-world-model/

76 105 114 https://arxiv.org/html/2512.04797v1
https://arxiv.org/html/2512.04797v1

77 118 https://www.reuters.com/business/videogame-stocks-slide-googles-ai-model-that-turns-prompts-into-playable-worlds-2026-01-30/
https://www.reuters.com/business/videogame-stocks-slide-googles-ai-model-that-turns-prompts-into-playable-worlds-2026-01-30/

87 120 https://blog.google/products-and-platforms/products/chrome/gemini-3-auto-browse/
https://blog.google/products-and-platforms/products/chrome/gemini-3-auto-browse/

92 https://docs.cloud.google.com/vertex-ai/generative-ai/docs/learn/model-versions
https://docs.cloud.google.com/vertex-ai/generative-ai/docs/learn/model-versions

116 https://gemini.google/release-notes/
https://gemini.google/release-notes/

122 https://support.google.com/gemini/answer/16316416?co=GENIE.Platform%3DDesktop&hl=en
https://support.google.com/gemini/answer/16316416?co=GENIE.Platform%3DDesktop&hl=en

125 https://developers.googleblog.com/introducing-veo-3-1-and-new-creative-capabilities-in-the-gemini-api/
https://developers.googleblog.com/introducing-veo-3-1-and-new-creative-capabilities-in-the-gemini-api/