Anthropic का Claude Mythos 5.1: साइबर-जीवविज्ञान में सबसे सक्षम AI, सीमित पहुंच और सुरक्षा-प्रयोग

समाचार-संदर्भ: Anthropic आधिकारिक पेज (4 नवंबर 2026 को देखा गया): Mythos 5.1 (1 सितंबर 2026), Cyber Verification Program विस्तार (6 अक्टूबर 2026), निर्यात-नियंत्रण छूट (1 जुलाई 2026)।

प्रकाशन: 4 नवंबर 2026

मुख्य परीक्षा (GS-III): विज्ञान-प्रौद्योगिकी — जनरेटिव AI, द्वि-उपयोग (dual-use) तकनीक, साइबर-सुरक्षा, AI शासन/निर्यात-नियंत्रण।

प्रिलिम्स: Anthropic, Claude मॉडल-परिवार, frontier AI, export controls, trusted access programs।



मौलिक प्रतीकात्मक चित्र: pcsmantra.com

खबरों में क्यों

Anthropic ने अपने सबसे सक्षम मॉडल-वर्ग **Claude Mythos 5.1** को **साइबर-सुरक्षा और जीवविज्ञान अनुसंधान** के लिए घोषित किया है (1 सितंबर 2026) — क्षमता इतनी उच्च कि पहुंच **केवल जाँचे गए (vetted) संगठनों** तक सीमित है। 6 अक्टूबर 2026 को **Cyber Verification Program** का विस्तार हुआ: सुरक्षा-दल Opus 5.5, Sonnet 5.5 और Mythos 5.1 पर उन्नत साइबर-क्षमताओं के लिए आवेदन कर सकते हैं। इससे पहले 1 जुलाई 2026 को अमेरिकी सरकार की मंजूरी के बाद **Mythos 5 पर निर्यात-नियंत्रण हटाए** गए थे। कीमत: **$10 प्रति मिलियन input / $50 प्रति मिलियन output टोकन**; डिफ़ॉल्ट **30-दिन डेटा-प्रतिधारण** सुरक्षा-निगरानी के लिए अनिवार्य।

Mythos क्यों अलग श्रेणी है

Mythos-वर्ग को Anthropic ने “frontier” क्षमता माना है: साइबर-सुरक्षा (कोड में भेद्यता खोजना) और जीवविज्ञान में इतना सक्षम कि **दुरुपयोग** भी संभव — साइबर-हमले या खतरनाक जैव-हथियार। इसलिए पहुंच **trusted access programs** से: साइबर-रक्षकों के लिए Cyber Verification Program और जीवन-विज्ञान शोधकर्ताओं के लिए Life Sciences Verification Program। कंपनी का **Claude Security** उत्पाद भी अब Mythos 5.1 पर चलता है।

Fable 5.1: वही मॉडल, सुरक्षा-कवच के साथ

व्यापक उपयोग के लिए Anthropic ने **Claude Fable 5.1** दिया — **अंतर्निहित मॉडल वही** (Mythos 5.1), परंतु सुरक्षा-प्रतिबंध सहित। नए सेफगार्ड अधिक सटीक हैं: Fable 5.1 स्रोत-कोड में **software vulnerabilities पहचान** सकता है, जबकि जीवविज्ञान-सुरक्षा अब सामान्य (benign) अनुरोधों पर **85% कम हस्तक्षेप** करती है (Fable 5 के लॉन्च-सेफगार्ड की तुलना में)। द्वि-उपयोग जैव/रसायन प्रश्न Opus मॉडलों को भेजे जाते हैं; **पेनेट्रेशन-टेस्टिंग, एक्सप्लॉइट-निर्माण और बाइनरी-आधारित vulnerability scanning पर रोक** बनी है। बेंचमार्क में Fable 5.1 की तुलना Fable 5, Opus 5 और GPT-5.6 Sol से — जिन कार्यों में सेफगार्ड हस्तक्षेप करते हैं, वहाँ अंक-हानि स्वीकार की गई है।

PCSMANTRA.COM

निर्यात-नियंत्रण की राजनीति

जून 2026 में Mythos 5 ("Mythos Preview" अद्यतन) कुछ जाँचे गए भागीदारों तक सीमित रहा; **12 जून को उपलब्धता बाधित** हुई और **1 जुलाई 2026** को अमेरिकी सरकार की मंजूरी पर कुछ US संगठनों के लिए **export controls हटाए** गए। 2 जून को **Project Glasswing** को 15+ देशों के ~150 नए संगठनों तक बढ़ाया गया। यह क्रम दिखाता है कि frontier AI अब **सेमीकंडक्टर-जैसी निर्यात-नियंत्रण नीति** का विषय बन चुका है — क्षमता-वितरण पर सरकारी अनुमति।

परीक्षा-दृष्टि: AI शासन के आयाम

(1) **द्वि-उपयोग समस्या** — वही मॉडल रक्षा और हमले दोनों में; समाधान: access-tiering, capability gating। (2) **सुरक्षा-बनाम-उपयोगिता trade-off** — 85% कम false-positive हस्तक्षेप का दावा: over-blocking भी विज्ञान को नुकसान। (3) **डेटा-गोपनीयता** — 30-दिन retention सुरक्षा-निगरानी के लिए, परंतु उपयोगकर्ता-सम्मति आवश्यक। (4) **भारत के लिए** — घरेलू frontier-क्षमता, AI-मिशन, और निर्यात-नियंत्रण व्यवस्था में आत्मनिर्भरता का प्रश्न; (5) **टोकन-मूल्य निर्धारण** — frontier मॉडलों की लागत-संरचना ($10/$50 प्रति मिलियन) बाजार-प्रवेश-बाधा।

घटनाक्रम

- **2 जून 2026:** Project Glasswing — 15+ देशों के ~150 नए संगठन
- **9 जून 2026:** Claude Mythos 5 — जाँचे गए भागीदारों तक सीमित
- **12 जून 2026:** Mythos 5 उपलब्धता बाधित
- **1 जुलाई 2026:** US सरकार की मंजूरी; निर्यात-नियंत्रण हटे
- **1 सितंबर 2026:** Claude Mythos 5.1 घोषित — साइबर व जीवविज्ञान में सुधार
- **6 अक्टूबर 2026:** Cyber Verification Program विस्तार (Opus 5.5, Sonnet 5.5, Mythos 5.1)

परीक्षा के लिए महत्त्व

मुख्य परीक्षा (GS-III): जनरेटिव AI और frontier मॉडल; द्वि-उपयोग तकनीक-नियंत्रण; साइबर-सुरक्षा; AI शासन, निर्यात-नियंत्रण और तकनीकी-सर्वभौमिकता।
प्रिलिम्स: Anthropic और Claude परिवार (Opus/Sonnet/Fable/Mythos); export controls; verification programs; टोकन-आधारित मूल्य-निर्धारण।

मुख्य परीक्षा अभ्यास प्रश्न

PCSMANTRA.COM

प्रश्न. "अत्यंत सक्षम AI मॉडलों का वितरण अब उनके निर्माण जितना ही बड़ा नीति-प्रश्न है।" Claude Mythos 5.1 के संदर्भ में टिप्पणी कीजिए। (250 शब्द, 15 अंक)

भूमिका में
Mythos 5.1 का "vetted access" मॉडल AI शासन की नई दिशा दर्शाता है।

मुख्य भाग में

द्वि-उपयोग जोखिम (साइबर/जैव); access-tiering और verification programs का मॉडल; निर्यात-नियंत्रण और सरकारी अनुमति; सुरक्षा-उपयोगिता trade-off (85% कम हस्तक्षेप); डेटा-प्रतिधारण और गोपनीयता।

निष्कर्ष में

भारत के लिए संदेश: घरेलू frontier-क्षमता + अपना सुरक्षित-पहुंच ढांचा — निर्यात-नियंत्रण युग में तकनीकी-आत्मनिर्भरता ही विकल्प-स्वतंत्रता है।

प्रिलिम्स परीक्षा अभ्यास प्रश्न

प्रश्न 1. Claude Mythos 5.1 के संदर्भ में निम्न कथनों पर विचार कीजिए:

कथन-I: इसकी पहुंच केवल जाँचे गए संगठनों तक सीमित है, क्योंकि यह साइबर-सुरक्षा और जीवविज्ञान में अत्यंत सक्षम है।
कथन-II: Claude Fable 5.1 और Mythos 5.1 का अंतर्निहित मॉडल एक ही है; Fable में अतिरिक्त सुरक्षा-प्रतिबंध हैं।

उपर्युक्त कथनों में से कौन-सा/से सही है/हैं?

(a) केवल कथन-I

(b) केवल कथन-II

(c) कथन-I और कथन-II दोनों

(d) न तो कथन-I, न कथन-II

उत्तर: (c) — दोनों सही: Mythos 5.1 vetted संगठनों तक सीमित है और Fable 5.1 वही मॉडल सुरक्षा-कवच के साथ है।

प्रश्न 2. AI संदर्भ में "द्वि-उपयोग (dual-use)" से क्या तात्पर्य है?

(a) एक ही मॉडल के दो भाषाओं में उपयोग

(b) एक ही क्षमता का उपयोग कल्याण और हानि — दोनों के लिए संभव होना

(c) दो कंपनियों द्वारा संयुक्त मालिकी

(d) मुफ्त और paid — दोनों संस्करणों की उपलब्धता

उत्तर: (b) — द्वि-उपयोग = वही तकनीक रक्षात्मक (भेद्यता-खोज) और आक्रामक (एक्सप्लॉइट-निर्माण) दोनों प्रयोग में आ सकती है — Mythos जैसे मॉडलों की सीमित पहुंच इसी कारण है।

PCSMANTRA.COM

स्रोत

1. Anthropic आधिकारिक पेज (4 नवंबर 2026 को देखा गया) — Mythos 5.1 घोषणा, Fable 5.1 सेफगार्ड, मूल्य-निर्धारण, verification programs, निर्यात-नियंत्रण समयरेखा
https://www.anthropic.com/claude/mythos