TROY TECHNICAL

AI & MACHINE LEARNING

WEEKLY REPORT

November 8 - November 14, 2026 | Issue 2026-W45

Deployable systems, permissions, infrastructure and exit readiness

15	10
ANALYZED ARTICLES	PRIORITY THEMES

Issued 2026-11-14 | English edition | fixed approved cover artwork



AI & Machine Learning: the boundary moved from isolated claims to qualification evidence

15 English articles were reviewed independently from the Japanese edition. Ten themes are retained for management decisions.

SIGNAL 01 | MODELS & DATA

Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests

Google DeepMind's Gemini Robotics 2.5 vision-language-action model achieved a 96% success rate in automotive parts-handling tests using Apptronik's Apollo 2 humanoid.

SIGNAL 02 | COMPUTE & DEVICES

Concept2Brain: AI-driven EEG simulation from images and text

Researchers have developed Concept2Brain, an open-access AI tool that generates synthetic electroencephalography (EEG) recordings based on photographs and written text.

SIGNAL 03 | INFRASTRUCTURE

AI Auscultation: Juntendo University's Heart Function Detection

Researchers from Juntendo University and AMI Inc. developed an AI model that detects left ventricular dysfunction using digital heart sounds and ECGs with an AUC of 0.904.

SIGNAL 04 | DEPLOYMENT

Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing

Anthropic has released Claude Fable 5.2 and Mythos 5.2, featuring a 20% reduction in API costs and up to 30% faster output speeds compared to version 5.1.

DECISION

Focus this week's management attention on deployable systems, permissions, infrastructure and exit readiness. Move only when the linked evidence can be reproduced under your own operating conditions.

THREE MANAGEMENT MOVES

1. Buyers: benchmark representative workflows and define permission, audit and shutdown requirements.
2. Suppliers: translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.
3. Executives: track integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EVIDENCE DISCIPLINE

FACT states what the collected English source reports. BUSINESS READ and ACTION are editorial interpretations. UNKNOWN lists information that the source file does not establish. A linked article is not treated as independent confirmation of every claim.

Numbers and evidence anchors that require conditions

96%
A12

SOURCE-BOUND CONDITION
Google DeepMind's Gemini Robotics 2.5 vision-language-action model achieved a 96% success rate in automotive parts-handling tests using Apptronik's Apollo 2 humanoid.

20%
A10

SOURCE-BOUND CONDITION
Anthropic has released Claude Fable 5.2 and Mythos 5.2, featuring a 20% reduction in API costs and up to 30% faster output speeds compared to version 5.1.

520 MHz
A09

SOURCE-BOUND CONDITION
STMicroelectronics has introduced the STM32H7P microcontroller, integrating a 520 MHz Arm Cortex-M7 CPU with a hardware signal processor (HSP) and optional NPU.

98.2%
A06

SOURCE-BOUND CONDITION
The model achieved 98.2% accuracy in classifying all 48 types of tRNA in E.

100%
A03

SOURCE-BOUND CONDITION
The system identified 100% of PTSD cases in a head-to-head test against a specialist, offering a scalable screening method to enable early preventive intervention.

HOW TO USE THESE NUMBERS
Each value is shown with the condition stated in the collected English article. Do not compare values across different test methods, device sizes, operating temperatures or commercial stages without normalizing those conditions.

Where the value chain moved

Each stage combines one leading theme and one supporting theme. The report treats handoffs as evidence transfers, not decorative arrows.

1

Models & data

P01/P06

Google DeepMind's Gemini Robotics 2.5 vision-language-action model achieved a 96% success rate in automotive parts-handling.

2

Compute & devices

P02/P07

Researchers have developed Concept2Brain, an open-access AI tool that generates synthetic electroencephalography (EEG).

3

Infrastructure

P03/P08

Researchers from Juntendo University and AMI Inc.

4

Deployment

P04/P09

Anthropic has released Claude Fable 5.2 and Mythos 5.2, featuring a 20% reduction in API costs and up to 30% faster output.

5

Governance

P05/P10

Axelera AI introduced Titania, a third-generation chiplet-based AI inference architecture.

EVIDENCE REQUIRED AT EACH HANDOFF

Models & data to Compute & devices	Transfer test conditions, acceptance criteria, failure data, owner and decision date. Recheck integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.
Compute & devices to Infrastructure	Transfer test conditions, acceptance criteria, failure data, owner and decision date. Recheck integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.
Infrastructure to Deployment	Transfer test conditions, acceptance criteria, failure data, owner and decision date. Recheck integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.
Deployment to Governance	Transfer test conditions, acceptance criteria, failure data, owner and decision date. Recheck integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

COMMON CONTROL POINT

No theme moves to a commercial commitment until a named owner has reproduced the relevant evidence and documented integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

Priority map: maturity and next decision locus



P01	Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests System / adoption | cue: deploying
P02	Concept2Brain: AI-driven EEG simulation from images and text System / adoption | cue: system
P03	AI Auscultation: Juntendo University's Heart Function Detection System / adoption | cue: system
P04	Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing Tech / process | cue: no system/business cue
P05	Axelera AI Titania: HPC-Class Inference Specs and Architecture System / adoption | cue: deployment
P06	AI Peer Review: Harvard Study Compares Claude Opus 4.8 vs Experts Tech / process | cue: no system/business cue
P07	STM32H7P MCU: Edge AI and DSP specs and 2027 production timeline System / adoption | cue: integrating
P08	tRNA Analysis: Deep Learning for Nanopore Signal Identification Tech / process | cue: no system/business cue
P09	PREVIA AI: PTSD Risk Prediction via Emergency Room Narratives System / adoption | cue: system
P10	Google Antigravity: Multi-Agent System for Math and Software System / adoption | cue: System

REPRODUCIBLE POSITIONING RULE

Horizontal position uses five discrete stages: Research, Prototype, Joint evaluation, Product adoption and Scale. The vertical bands classify the next decision locus: Tech / process, System / adoption, or Business / rules (including supply systems). Article-title cues take precedence; the overview is used only when the title is silent. These are categories, not impact ranks. Bubble diameter reflects the editorial score.

FACT is source-bound. BUSINESS READ, ACTION and UNKNOWN are explicit editorial layers.

P01 Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests

A12 | Models & data | Product adoption

2026-11-11

FACT

Google DeepMind's Gemini Robotics 2.5 vision-language-action model achieved a 96% success rate in automotive parts-handling tests using Apptronik's Apollo 2 humanoid. The model demonstrated success rates 30% to 100% higher than previous baselines and completed tasks 20% to 30% faster. This indicates a scalable foundation for deploying humanoid robots in industrial and commercial environments.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P01.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 21/25 | MATURITY 4/5 | DECISION LOCUS: SYSTEM / ADOPTION

P02 Concept2Brain: AI-driven EEG simulation from images and text

A04 | Compute & devices | Pilot

2026-11-13

FACT

Researchers have developed Concept2Brain, an open-access AI tool that generates synthetic electroencephalography (EEG) recordings based on photographs and written text. By simulating human brain activity during visual and emotional perception, the system allows researchers to conduct preliminary experiments on computers before recruiting human volunteers.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P02.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 2/5 | DECISION LOCUS: SYSTEM / ADOPTION

FACT is source-bound. BUSINESS READ, ACTION and UNKNOWN are explicit editorial layers.

P03 AI Auscultation: Juntendo University's Heart Function Detection

A07 | Infrastructure | Pilot

2026-11-12

FACT

Researchers from Juntendo University and AMI Inc. developed an AI model that detects left ventricular dysfunction using digital heart sounds and ECGs with an AUC of 0.904. The system outperformed the diagnostic accuracy of cardiology specialists in external validation, offering a scalable method for early heart failure screening.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P03.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 2/5 | DECISION LOCUS: SYSTEM / ADOPTION

P04 Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing

A10 | Deployment | Research

2026-11-10

FACT

Anthropic has released Claude Fable 5.2 and Mythos 5.2, featuring a 20% reduction in API costs and up to 30% faster output speeds compared to version 5.1. These high-end models differ primarily by their safeguard configurations, targeting general use and specialized research respectively.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P04.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 1/5 | DECISION LOCUS: TECH / PROCESS

FACT is source-bound. BUSINESS READ, ACTION and UNKNOWN are explicit editorial layers.

P05 Axelera AI Titania: HPC-Class Inference Specs and Architecture

A08 | Governance | Architecture disclosed

2026-11-12

FACT

Axelera AI introduced Titania, a third-generation chiplet-based AI inference architecture. The company expects up to seven times more throughput per watt and up to twenty times better total cost of ownership versus its chosen high-end server comparison. These are company performance expectations; independent workload-level results, pricing and deployment economics are not yet established.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P05.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 2/5 | DECISION LOCUS: SYSTEM / ADOPTION

P06 AI Peer Review: Harvard Study Compares Claude Opus 4.8 vs Experts

A05 | Models & data | Pilot

2026-11-12

FACT

AI-generated peer reviews outscored high-paid human experts in blind comparisons, showing a lower rate of incorrect claims and higher quality ratings. The findings suggest a potential shift in scientific journal evaluation processes, though they raise concerns regarding the uniformity of scientific thinking.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P06.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 2/5 | DECISION LOCUS: TECH / PROCESS

FACT is source-bound. BUSINESS READ, ACTION and UNKNOWN are explicit editorial layers.

P07 STM32H7P MCU: Edge AI and DSP specs and 2027 production timeline

A09 | Compute & devices | Product adoption

2026-11-10

FACT

STMicroelectronics has introduced the STM32H7P microcontroller, integrating a 520 MHz Arm Cortex-M7 CPU with a hardware signal processor (HSP) and optional NPU. The chip achieves up to 85 GOPS of AI performance, significantly accelerating inference for industrial power and monitoring applications.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P07.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 20/25 | MATURITY 4/5 | DECISION LOCUS: SYSTEM / ADOPTION

P08 tRNA Analysis: Deep Learning for Nanopore Signal Identification

A06 | Infrastructure | Pilot

2026-11-12

FACT

Researchers from the University of Tokyo developed a deep learning method to identify tRNA types, quantities, and modifications directly from nanopore current signals. The model achieved 98.2% accuracy in classifying all 48 types of tRNA in E. coli, providing a foundation for studying protein synthesis and disease.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P08.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 19/25 | MATURITY 2/5 | DECISION LOCUS: TECH / PROCESS

FACT is source-bound. BUSINESS READ, ACTION and UNKNOWN are explicit editorial layers.

P09 PREVIA AI: PTSD Risk Prediction via Emergency Room Narratives

A03 | Deployment | Research

2026-11-09

FACT

Researchers at NYU Langone Health developed PREVIA, an AI tool that predicts PTSD risk by analyzing patient narratives during emergency department (ED) intake. The system identified 100% of PTSD cases in a head-to-head test against a specialist, offering a scalable screening method to enable early preventive intervention.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P09.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 19/25 | MATURITY 1/5 | DECISION LOCUS: SYSTEM / ADOPTION

P10 Google Antigravity: Multi-Agent System for Math and Software

A13 | Governance | Research

2026-11-14

FACT

Google's Antigravity multi-agent system, powered by Gemini models, has formalized the Weil conjectures and rewritten critical software components in Rust. The framework utilizes specialized agents to achieve high accuracy in molecular structure identification and significant performance gains in video decoding.

WHY IT MATTERS

The decision relevance is deployable systems, permissions, infrastructure and exit readiness. The signal should be tested at the application and operating-system level, not accepted from a headline alone.

BUSINESS READ

For operating companies, compare the reported change with current qualification, sourcing and product-roadmap assumptions. For suppliers, translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence.

ACTION

Within 30 days, benchmark representative workflows and define permission, audit and shutdown requirements; record the evidence and owner against P10.

UNKNOWN

Still unproven or incompletely stated: integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.

EDITORIAL SCORE 19/25 | MATURITY 1/5 | DECISION LOCUS: SYSTEM / ADOPTION

Playbook and 0-5 year roadmap

STAKEHOLDER	NOW | 0-30 DAYS	NEXT | 1-12 MONTHS	LATER | 1-5 YEARS
Operating companies	benchmark representative workflows and define permission, audit and shutdown requirements	Run production-like qualification with explicit acceptance and stop criteria.	Build a portfolio and architecture that can absorb technology and supplier changes.
Materials, equipment and technology suppliers	translate model or hardware capability into workload-level latency, energy, reliability and lifecycle evidence	Create shared reliability, process-window and failure-analysis data with customers.	Standardize interfaces, traceability, lifecycle support and recovery services.
Trading, investment and legal teams	Map integration timing, operating cost, neutrality, permissions and measurable workflow outcomes.	Contract for capacity, quality, change notice, liability, alternatives and exit.	Diversify regional, technical and customer concentration risk.

ROADMAP

NOW | 0-30 DAYS
Evidence ledger
Owner: Strategy + quality
Deliverable: claim, source, condition, owner

NOW | 0-30 DAYS
Risk screen
Owner: Procurement + legal
Deliverable: unknowns, alternatives, stop

NEXT | 1-12 MONTHS
Joint qualification
Owner: R&D + supplier
Deliverable: process window and failure data

NEXT | 1-12 MONTHS
Commercial gate
Owner: Business + finance
Deliverable: unit economics and capacity

LATER | 1-5 YEARS
Platform strategy
Owner: Executive team
Deliverable: portfolio and interface roadmap

LATER | 1-5 YEARS
Service layer
Owner: Operations + channel
Deliverable: monitoring, maintenance and recovery

GO / NO-GO GATES

GATE	OWNER	REQUIRED EVIDENCE	NO-GO CONDITION
G1 Evidence	**Quality**	Source, test conditions and reproducibility	No traceable evidence
G2 Integration	**R&D**	Interface, process window and failure analysis	Cannot meet system conditions
G3 Commercial	**Business**	Capacity, total cost, contract and alternatives	Economics or supply cannot be supported
G4 Lifecycle	**Operations**	Monitoring, maintenance, change notice and exit	No accountable operating plan

ANALYZED ARTICLES | ITEMS 01-15

Every title and URL is displayed in full. URLs wrap without shortening and remain clickable. A verified-reference label means the translated HTML did not retain its original source URL.

ID	FULL ARTICLE TITLE	FULL SOURCE / VERIFIED REFERENCE URL - CLICKABLE
A01	WorldParticle: Unified AI Model for Multi-Substance Simulation	https://news.mit.edu/2026/single-ai-model-simulates-many-substances-behavior-1109
A02	AlphaGenome 2: DeepMind's AI for cell-specific genome mapping	https://www.nature.com/articles/d41586-026-03673-0
A03	PREVIA AI: PTSD Risk Prediction via Emergency Room Narratives	https://www.prnewswire.com/news-releases/ai-predicts-ptsd-risk-based-on-stories-about-what-brought-patients-to-the-emergency-room-302926212.html
A04	Concept2Brain: AI-driven EEG simulation from images and text	https://www.psypost.org/scientists-built-an-ai-called-concept2brain-that-simulates-human-brain-waves-from-pictures-and-text/
A05	AI Peer Review: Harvard Study Compares Claude Opus 4.8 vs Experts	https://www.psypost.org/ai-written-peer-reviews-outscored-top-human-experts-in-blind-comparisons-harvard-working-paper-finds/
A06	tRNA Analysis: Deep Learning for Nanopore Signal Identification	https://www.nikkei.com/article/DGXZRSP715197_S6A111C2000000/
A07	AI Auscultation: Juntendo University's Heart Function Detection	https://www.jiji.com/jc/article?k=000000924.000021495&g=prt
A08	Axelera AI Titania: HPC-Class Inference Specs and Architecture	https://finance.yahoo.com/technology/ai/articles/axelera-ai-unveils-titania-architecture-140000512.html
A09	STM32H7P MCU: Edge AI and DSP specs and 2027 production timeline	https://biz.chosun.com/en/en-it/2026/11/10/L6TQI35RDJCBZHLCGXA2R4CKYY/
A10	Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing	https://www.itmedia.co.jp/news/article/2611/10/2000003042/
A11	Meta Muse Ember: Low-cost AI model specs and performance vs Spark	https://www.itmedia.co.jp/news/article/2611/10/2000003044/
A12	Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests	https://theaiinsider.tech/2026/11/11/google-deepmind-unveils-gemini-robotics-2-5-reports-96-success-in-mercedes-benz-factory-test/
A13	Google Antigravity: Multi-Agent System for Math and Software	https://interestingengineering.com/ai-robotics/google-antigravity-multi-agent-research
A14	NTT World Model: Physical AI for Robot Failure Avoidance	https://news.mynavi.jp/techplus/article/20261111-5296066/
A15	Ripple-65B Preview: Ricursive Intelligence's Open Model Specs	https://startuprecipe.co.kr/archives/tech/5840663

AI_MachineLearning — Selected Articles

Date: 2026-11-15

Articles: 15

Table of Contents

#01 WorldParticle: Unified AI Model for Multi-Substance Simulation

#02 AlphaGenome 2: DeepMind's AI for cell-specific genome mapping

#03 PREVIA AI: PTSD Risk Prediction via Emergency Room Narratives

#04 Concept2Brain: AI-driven EEG simulation from images and text

#05 AI Peer Review: Harvard Study Compares Claude Opus 4.8 vs Experts

#06 tRNA Analysis: Deep Learning for Nanopore Signal Identification

#07 AI Auscultation: Juntendo University's Heart Function Detection

#08 Axelera AI Titania: HPC-Class Inference Specs and Architecture

#09 STM32H7P MCU: Edge AI and DSP specs and 2027 production timeline

#10 Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing

#11 Meta Muse Ember: Low-cost AI model specs and performance vs Spark

#12 Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests

#13 Google Antigravity: Multi-Agent System for Math and Software

#14 NTT World Model: Physical AI for Robot Failure Avoidance

#15 Ripple-65B Preview: Ricursive Intelligence's Open Model Specs

#01 WorldParticle: Unified AI Model for Multi-Substance Simulation

Published November 09, 2026 MIT News USA



OVERVIEW

Researchers from MIT and the University of British Columbia have developed WorldParticle, a single deep-learning system capable of simulating diverse physical phenomena from sliding sand to folding proteins. By replacing material-specific equations with a unified neural network architecture, the system enables the simulation of complex interactions between different physical regimes.

IN DEPTH

Key Findings

WorldParticle is a unified artificial intelligence architecture that simulates the behavior of various substances without requiring separate, hand-coded solvers for each material. Unlike traditional simulators that rely on specific equations for flowing water or protein folding, this deep-learning system can model multiple physical phenomena within a single framework. This allows for the simulation of coupled processes, such as fluid sloshing against a deformable solid, which typically requires two separate simulators.

Technical Details

The system utilizes a Lagrangian representation, breaking objects into thousands or millions of tiny three-dimensional particles. The process operates as follows:

- **Input:** The neural network takes the 3D positions and physical properties of particles as inputs.
- **Prediction:** It predicts the positions and properties for the next frame in one-second increments, effectively predicting particle movement one step at a time.
- **Force Integration:** WorldParticle separates external forces (such as impulse forces from an ocean wave hitting a seawall) before the simulation step and integrates them back afterward to correct particle predictions.

By learning underlying dynamics rather than being programmed with fixed equations, the model can adapt to forces, materials, and environments not encountered during its training phase.

Background & Context

The research was a collaborative effort involving MIT, the University of British Columbia (UBC), Georgia Tech, Inria, Meta, and several other universities. The team drew inspiration from "world models," AI systems that learn physical rules to predict environmental changes over time. The work will be presented at the ACM SIGGRAPH Conference and Exhibition on Computer Graphics and Interactive Techniques in Asia.

Significance & Outlook

The researchers describe this as a first step toward a "GPT moment for simulation," aiming for a foundation model that works out of the box for any scenario. Potential applications include studying how granular materials like soil move when touched by soft robots or analyzing how oil impacts elastic valves in storage tanks. Long-term, the system is intended to help scientists discover new materials and explore physical processes that are currently inaccessible via existing simulation methods.

Source: https://news.mit.edu/2026/single-ai-model-simulates-many-substances-behavior-1109

Collected: November 14, 2026 | Automated Research System (AI)

#02 AlphaGenome 2: DeepMind's AI for cell-specific genome mapping

Published November 11, 2026 Nature UK



OVERVIEW

Google DeepMind has developed AlphaGenome 2, an AI model capable of predicting the molecular effects of non-coding DNA variants across hundreds of human and mouse cell types. The model replicated a complex genetic discovery regarding inflammatory bowel disease in 30 minutes that previously required 1.5 years of laboratory experiments. This advancement enables faster identification of genetic variants underlying diseases by providing critical cell-type specificity.

IN DEPTH

Key Findings

Google DeepMind's AlphaGenome 2 can predict how "non-coding" DNA sequences affect the expression of other genes across 637 types of human cells and 311 kinds of mouse cells. In a validation case, the model replicated findings related to a DNA variant in the ETS2 gene—a driver of inflammatory bowel disease (IBD)—in approximately 30 minutes. This specific discovery had previously taken researchers at the Francis Crick Institute a year and a half of experimental work using gene-editing and molecular-biology techniques.

Technical Details

AlphaGenome 2 improves upon its predecessor, AlphaGenome 1, by addressing the lack of cell-type specificity. While AlphaGenome 1 was trained on "bulk" samples of heterogeneous cells, AlphaGenome 2 was trained on measurements from more than 100 million cells across 68 tissues, utilizing single-cell gene activity data. This approach allows the model to provide a more fine-grained representation of cellular differences. Additionally, the model is flexible; researchers can input new single-cell gene-activity data, such as data from cells in a disease state, to generate specific gene-variant predictions.

Background & Context

Non-coding DNA sequences comprise 98% of the human genome and are among the most difficult to decipher. Many of these sequences, including gene enhancers, are active only in specific cell types. Previously, Google DeepMind released the AlphaGenome Atlas, which provided predictions for nine billion possible single-letter DNA changes to the human genome, but these predictions did not specify which cell types—such as pancreatic β-cells or B cells—were affected.

Significance & Outlook

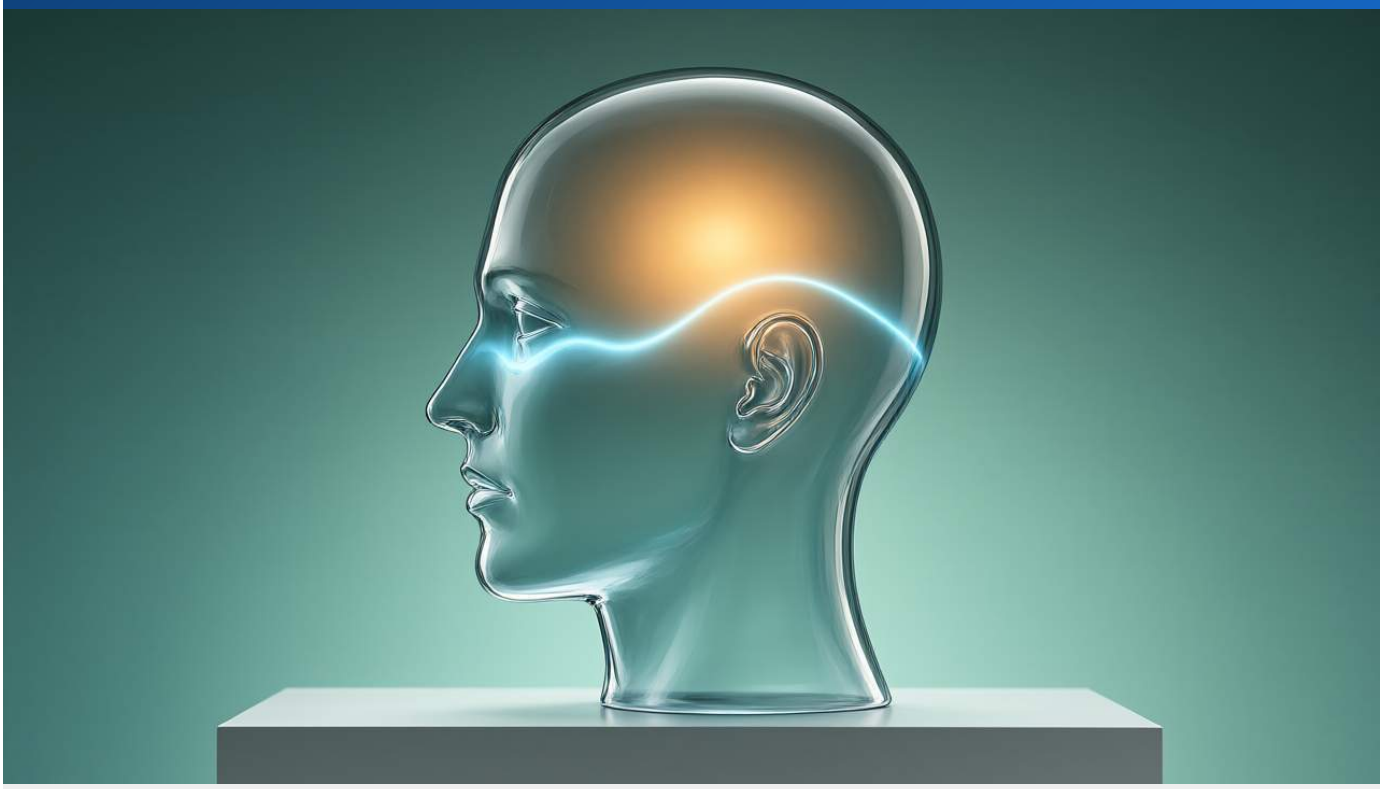
The ability to predict variant effects in diverse cells is critical for understanding genetic diseases. By eliminating the need for immediate physical experimentation (described as "lifting a pipette"), AlphaGenome 2 allows researchers to identify disease-underlying variants more rapidly. This capability streamlines the process of mapping how variants in "gene deserts" influence protein-coding sequences and disease pathology.

Source: https://www.nature.com/articles/d41586-026-03673-0

Collected: November 14, 2026 | Automated Research System (AI)

#03 PREVIA AI: PTSD Risk Prediction via Emergency Room Narratives

Published November 09, 2026 PR Newswire（NYU Langone Health） USA



OVERVIEW

Researchers at NYU Langone Health developed PREVIA, an AI tool that predicts PTSD risk by analyzing patient narratives during emergency department (ED) intake. The system identified 100% of PTSD cases in a head-to-head test against a specialist, offering a scalable screening method to enable early preventive intervention.

IN DEPTH

Key Findings

The AI tool PREVIA correctly identified more than 95.4 percent of at-risk patients in a study involving 145 emergency department (ED) patients. In a direct comparison, PREVIA flagged 100 percent of 30 cases that later developed PTSD, outperforming an ED clinician specializing in PTSD, who correctly identified 89 percent of those cases. The system utilizes spontaneous answers to the routine question, "What brought you to the ED?", to forecast risk one month later.

Technical Details

PREVIA is built on open-source models without requiring additional training or fine-tuning. It employs a three-layer pipeline to synthesize risk factors:

- **First Layer:** Scans narratives for peritraumatic symptoms (e.g., amnesia, loss of control) and early ED symptoms (e.g., feeling disconnected from the body).
- **Second Layer:** Weighs risk factors, such as lower socioeconomic status and pre-existing mental health issues, against protective factors like social support and coping skills. This layer utilizes different models for specific strengths; for example, Llama emphasizes risk signs while Mistral highlights protective factors.
- **Third Layer:** Synthesizes assessments into a structured summary to prevent any single model from dominating the result.

The final output combines this summary with demographic data (age, sex), trauma type, and PTSD prevalence rates to produce a risk score with a plain-language explanation. Results were validated using the Clinician-Administered PTSD Scale for DSM-5 (CAPS-5).

Background & Context

Approximately 43 million patients are admitted to US emergency departments annually for physical injuries, with over a quarter developing PTSD. While validated tools like the Posttraumatic Adjustment Scale exist, only 28 percent of EDs routinely screen for PTSD, and only two-thirds of those use validated tools. This gap is largely due to time constraints, as clinicians typically have only 15–20 minutes per patient.

Significance & Outlook

Because many patients lose contact with hospitals after the first month, the ED represents a critical window for early detection. By providing an automated screener that requires no additional intake time, PREVIA enables timely preventive interventions. The use of open-source models allows other hospitals to test the tool independently without proprietary software.

Source: https://www.prnewswire.com/news-releases/ai-predicts-ptsd-risk-based-on-stories-about-what-brought-patients-to-the-emergency-room-302926212.html

Collected: November 14, 2026 | Automated Research System (AI)

#04 Concept2Brain: AI-driven EEG simulation from images and text

Published November 13, 2026 PsyPost USA



OVERVIEW

Researchers have developed Concept2Brain, an open-access AI tool that generates synthetic electroencephalography (EEG) recordings based on photographs and written text. By simulating human brain activity during visual and emotional perception, the system allows researchers to conduct preliminary experiments on computers before recruiting human volunteers.

IN DEPTH

Key Findings

Concept2Brain is a generative AI tool designed to predict how human brain waves respond to specific visual or textual stimuli. Published in Nature Communications, the research demonstrates that the platform can produce synthetic EEG recordings that resemble real human brain activity. The system effectively bridges the gap between conceptual meaning and scalp electrical activity, allowing for the simulation of brain-wave experiments without the immediate need for costly and time-consuming human data collection.

Technical Details

The architecture consists of two linked deep neural networks:

- **Conditional Variational Autoencoder:** This network compresses 129-channel EEG recordings into a compact summary of 256 numbers. It includes a participant identifier to preserve individual differences in brain-wave shapes.
- **Translation Network:** Built around OpenAI's CLIP system, this network maps images and text into a shared 512-number conceptual space. A three-layer neural network then translates these conceptual numbers into the 256-number brain-wave summary.

To train the model, researchers used EEG data from 88 university volunteers across two experiments. Participants viewed 360 unique photographs—divided equally among pleasant, neutral, and unpleasant scenes—recorded via 128 scalp sensors and one reference sensor at 500 samples per second. After filtering noise, the team utilized 13,640 clean 1.5-second trials.

Background & Context

While AI has previously been used for "decoding" (reconstructing meaning from brain scans), Concept2Brain focuses on "encoding"—predicting electrical waves from a starting image or phrase. This builds on previous research, including a 2023 multimodal project cataloging 1,854 object categories and a 2008 study linking emotional arousal to the magnitude of electrical brain responses.

Significance & Outlook

Lead author Alejandro Santos-Mayo states that the tool provides a non-linear, multivariate approach to predicting brain data. By offering the model's code, trained weights, and a public website, the researchers aim to provide a free, open-access resource for simulating EEG activity and deepening the understanding of neurophysiological components.

Source: https://www.psypost.org/scientists-built-an-ai-called-concept2brain-that-simulates-human-brain-waves-from-pictures-and-text/

Collected: November 14, 2026 | Automated Research System (AI)

#05 AI Peer Review: Harvard Study Compares Claude Opus 4.8 vs Experts

Published November 12, 2026 PsyPost USA



OVERVIEW

AI-generated peer reviews outscored high-paid human experts in blind comparisons, showing a lower rate of incorrect claims and higher quality ratings. The findings suggest a potential shift in scientific journal evaluation processes, though they raise concerns regarding the uniformity of scientific thinking.

IN DEPTH

Key Findings

AI-written peer reviews outperformed top human experts in blind comparisons, according to a Harvard University working paper. In evaluations conducted by both a separate AI model and former journal editors, AI reviews won the large majority of matchups. Notably, AI-generated reviews contained approximately one-third the rate of incorrect claims compared to those written by humans.

Technical Details

Researchers compared human reviews from rapidPeer—a service where experts are paid well to provide 1.5 to 2 pages of feedback—against reviews generated by Anthropic’s Claude Opus 4.8. The dataset included 124 scholarly manuscripts and 134 human reviews from 2019 to 2026 across mathematical, statistical, and qualitative fields.

The AI was tested under two conditions: a "matched" condition using only the paper and reviewer instructions, and a "tuned" condition using a custom prompt and internet search access, averaging 12.4 searches per review. OpenAI’s GPT-5.5 served as the AI judge, ranking pairs across seven areas: theory and idea, methods, weaknesses, framing, communication, figures and tables, and sources. In a check providing the full manuscript, the AI review was preferred in at least 132 of 134 pairs.

Background & Context

Traditional peer review faces challenges including delays, recruitment difficulties, and human biases toward prestigious institutions. While previous 2026 studies yielded conflicting results—one showing human reviewers provided more varied points at BMJ Open, and another showing AI reviews were rated higher in a cardiology journal—this study specifically tested AI against experts performing at their best rather than unpaid referees.

Significance & Outlook

The results indicate that advanced AI can generate more thorough and constructive feedback than human experts operating under ideal conditions. However, the researchers note that widespread reliance on these tools could potentially make scientific thinking more uniform. The working paper has not yet been peer reviewed.

Source: https://www.psypost.org/ai-written-peer-reviews-outscored-top-human-experts-in-blind-comparisons-harvard-working-paper-finds/

Collected: November 14, 2026 | Automated Research System (AI)

#06 tRNA Analysis: Deep Learning for Nanopore Signal Identification

Published November 12, 2026　日本経済新聞（東京大学プレスリリース）　Japan



OVERVIEW

Researchers from the University of Tokyo developed a deep learning method to identify tRNA types, quantities, and modifications directly from nanopore current signals. The model achieved 98.2% accuracy in classifying all 48 types of tRNA in E. coli, providing a foundation for studying protein synthesis and disease.

IN DEPTH

Key Findings

A research group led by Special Lecturer Hiroaki Ueda and Professor Tsutomu Suzuki from the University of Tokyo has developed a technology to integratedly analyze the type, abundance, and modification state of transfer RNA (tRNA). Using a deep learning model called tyCooNN, the team created a "nano tRNA classifier" that identifies tRNA types directly from current signals measured by a nanopore sequencer. In tests using E. coli, the model achieved a classification accuracy of 98.2% across all 48 tRNA types.

Technical Details

The developed method bypasses "basecalling"—the process of converting current signals into base sequences—and instead analyzes the raw current patterns. To address the challenge of quantification, the researchers established an absolute quantification method. By using a standard sample consisting of purified tRNAs mixed in equal molar amounts, they were able to correct for biases in reading efficiency specific to each tRNA species. This allowed for the comparison of different tRNA abundances using a common baseline.

Additionally, the team developed a technique to determine the presence or absence of specific RNA modifications at the single-molecule level. By training the model on signals from tRNAs with and without specific modifications, they demonstrated the ability to estimate the modification rate for each tRNA species.

Background & Context

tRNA molecules correspond genetic codes in messenger RNA (mRNA) to amino acids. Their abundance and modification states influence the speed and accuracy of protein synthesis, mRNA stability, and protein quality maintenance. However, because tRNAs have highly similar sequences and contain numerous chemical modifications, traditional sequencing analysis has struggled with accurate identification and quantification.

Significance & Outlook

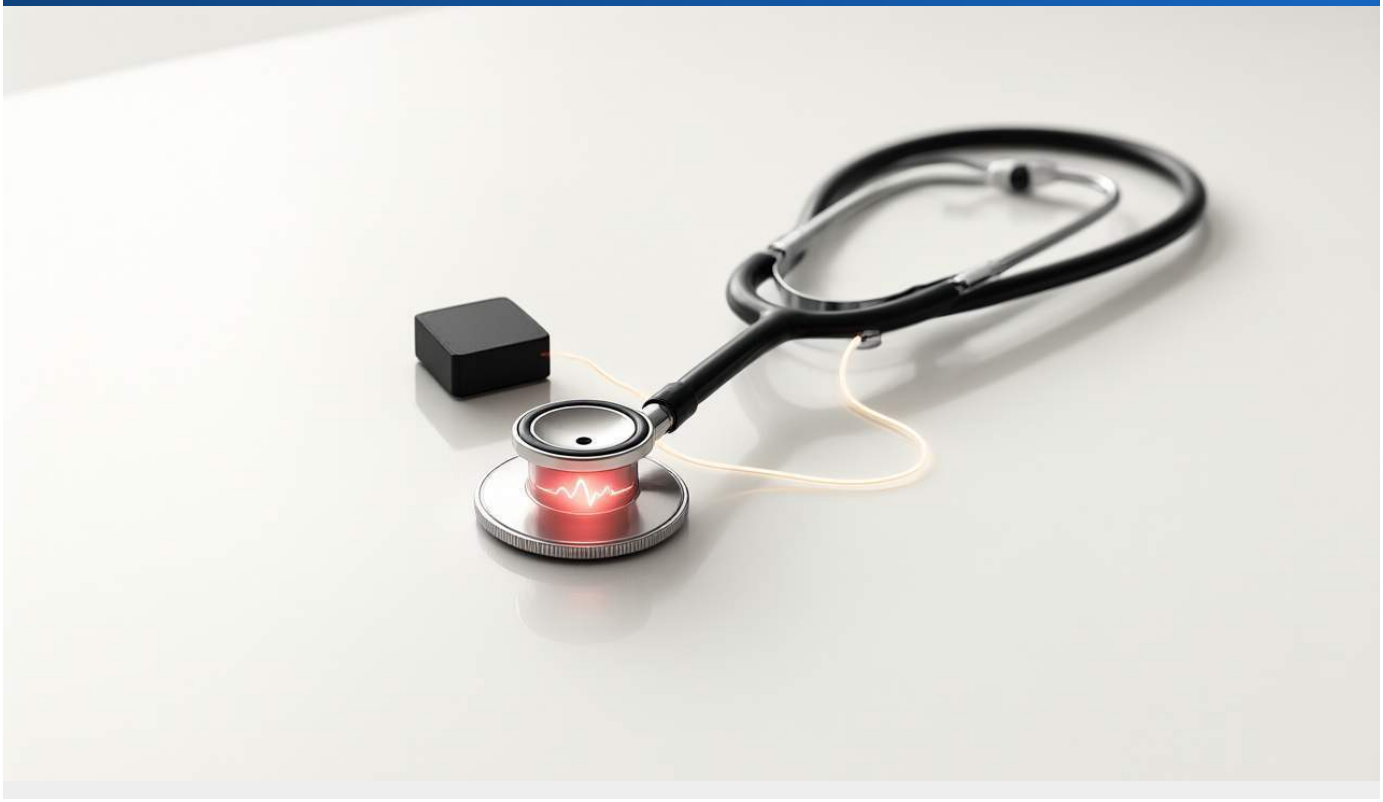
The researchers applied this method to E. coli, capturing changes in tRNA composition associated with nutritional conditions and growth stages. This integrated analysis of tRNA quantity and modification status serves as a technical basis for investigating cellular protein synthesis and its relationship to diseases. The findings were published in Nature Communications on November 9, 2026.

Source: https://www.nikkei.com/article/DGXZRSP715197_S6A111C2000000/

Collected: November 14, 2026 | Automated Research System (AI)

#07 AI Auscultation: Juntendo University's Heart Function Detection

Published November 12, 2026　時事ドットコム（順天堂大学プレスリリース）　Japan



OVERVIEW

Researchers from Juntendo University and AMI Inc. developed an AI model that detects left ventricular dysfunction using digital heart sounds and ECGs with an AUC of 0.904. The system outperformed the diagnostic accuracy of cardiology specialists in external validation, offering a scalable method for early heart failure screening.

IN DEPTH

Key Findings

An AI-driven auscultation method has demonstrated high precision in detecting left ventricular dysfunction, defined as a left ventricular ejection fraction (LVEF) of 40% or less. In an external validation study involving 964 patients across six facilities, the AI achieved an Area Under the Curve (AUC) of 0.904, a sensitivity of 86.3%, and a negative predictive value of 98.5%. When compared directly against seven cardiology specialists, the AI's accuracy was 82.0%, significantly higher than the 62.0% accuracy achieved by the specialists' majority vote.

Technical Details

The system utilizes the "Cardio-EGG" digital stethoscope from AMI Inc. to simultaneously record single-lead ECGs and heart sounds from four standard chest locations for eight seconds each. The AI architecture integrates a Transformer model, designed to capture temporal sequences in audio and electrical data, with a Convolutional Neural Network (CNN) for extracting local waveform features. The model was trained on a dataset of 2,359 individuals (average age 69.7) before being validated on the separate cohort of 964 individuals (average age 68.6). The AI outperformed models using only heart sounds (AUC 0.832), only ECGs (AUC 0.804), or statistical models combining clinical data and BNP blood tests (AUC 0.719–0.848).

Background and Context

Left ventricular dysfunction often lacks obvious symptoms, leading to delayed diagnosis of heart failure. While echocardiography is the gold standard for evaluation, it requires specialized equipment and technical expertise. Traditional auscultation is simple but subjective, with results varying based on the physician's experience. This research sought to provide an objective, AI-enhanced alternative to identify patients who require urgent echocardiography.

Significance and Outlook

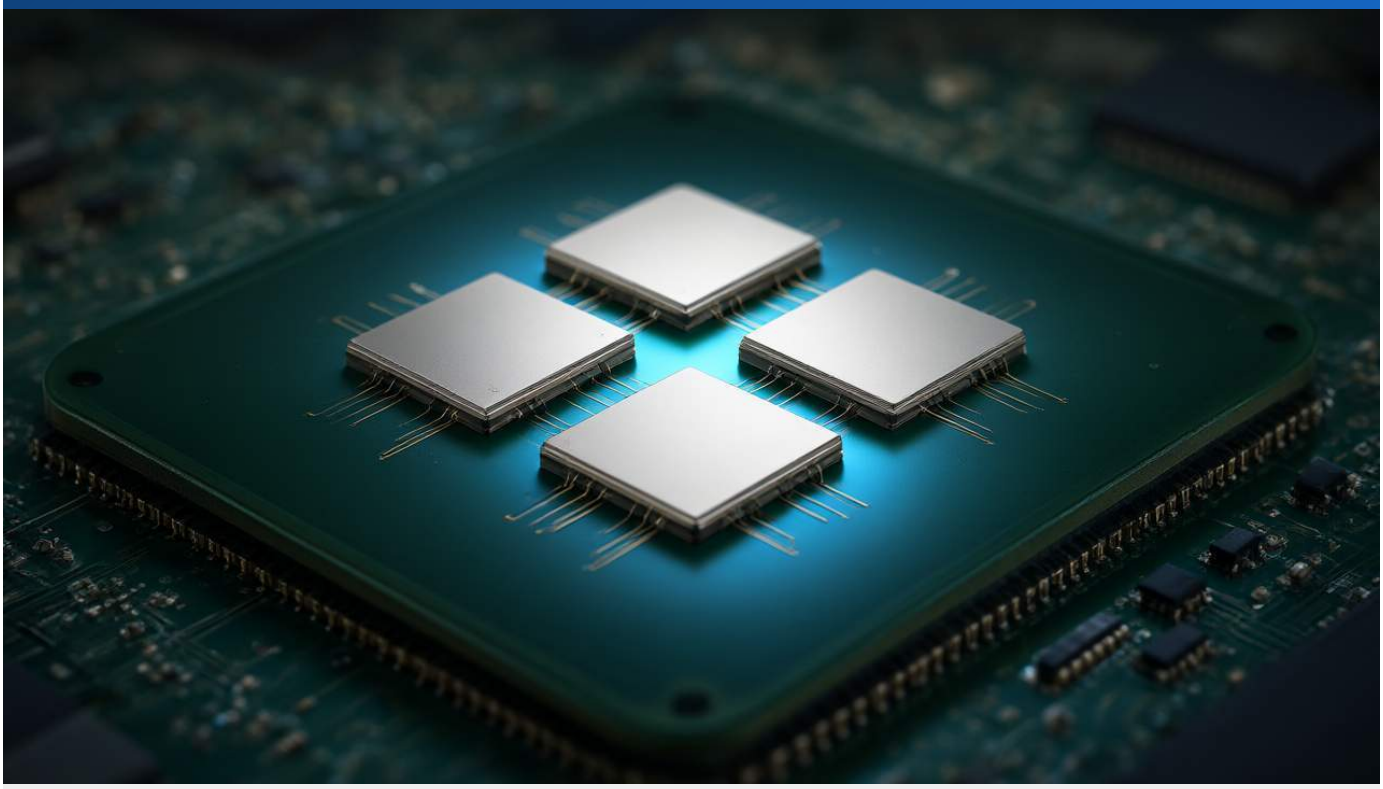
Because digital auscultation is fast and requires no heavy infrastructure, it can be deployed in medical settings where immediate echocardiography is unavailable. While the current study focused on patients already deemed likely to need echocardiography, future research will evaluate the system's utility in general health screenings and its impact on patient prognosis and clinical efficiency.

Source: https://www.jiji.com/jc/article?k=000000924.000021495&g=prt

Collected: November 14, 2026 | Automated Research System (AI)

#08 Axelera AI Titania: HPC-Class Inference Specs and Architecture

Published November 12, 2026　Business Wire（Yahoo Finance掲載）　Netherlands



OVERVIEW

Axelera AI has introduced Titania, a third-generation chiplet-based architecture designed to deliver data center-grade AI performance within existing air-cooled rack constraints. The platform targets a $110B underserved market by offering up to 7x throughput per watt and 20x better TCO compared to current high-end servers.

IN DEPTH

Key Findings

The Titania architecture enables high-performance AI inference without the need for 100-kilowatt-plus liquid-cooled racks. It is designed to fit into existing enterprise and lab racks, providing a cost and power budget lower than an RTX6000 server. The company reports confirmed commitments for Titania exceeding $300 million.

Technical Details

Titania utilizes a chiplet design based on Digital In-Memory Computing (D-IMC) and incorporates programmable RISC-V compute and high-capacity LPDDR6 memory. By using LPDDR6 instead of HBM, each package achieves 128 GB of capacity with lower power consumption and avoids HBM supply chain dependencies.

- **Form Factors:** Available as standard PCIe cards and OCP-compliant UBB baseboards with eight Titania OAM modules.
- **Reference Server (AX8):** A 4U air-cooled configuration providing 20 PFLOPS of FP8 compute, 1 TB of accelerator memory, and 19.6 TB/s memory bandwidth.
- **Power Consumption:** Typical power draw of 5-6kW for the AX8 server.
- **Model Capacity:** A single server can hold a 1.6 trillion-parameter model at FP4. Two servers fit in a legacy 15 kW rack, and five fit in a 30 kW air-cooled rack.
- **Software Support:** The Voyager toolchain and AxeleraScript (AxScript) support tensor, expert, and pipeline parallelism, speculative decoding, and KV cache offloading.

Background and Context

Titania is the third generation of Axelera AI's technology, following the Metis and Europa platforms. While Metis focused on energy efficiency and Europa extended the architecture into enterprise AI, Titania expands the product line into large-scale data center inference. It is specifically designed for interactive and variable workloads in sovereign clouds, telcos, and AI factories.

Significance and Outlook

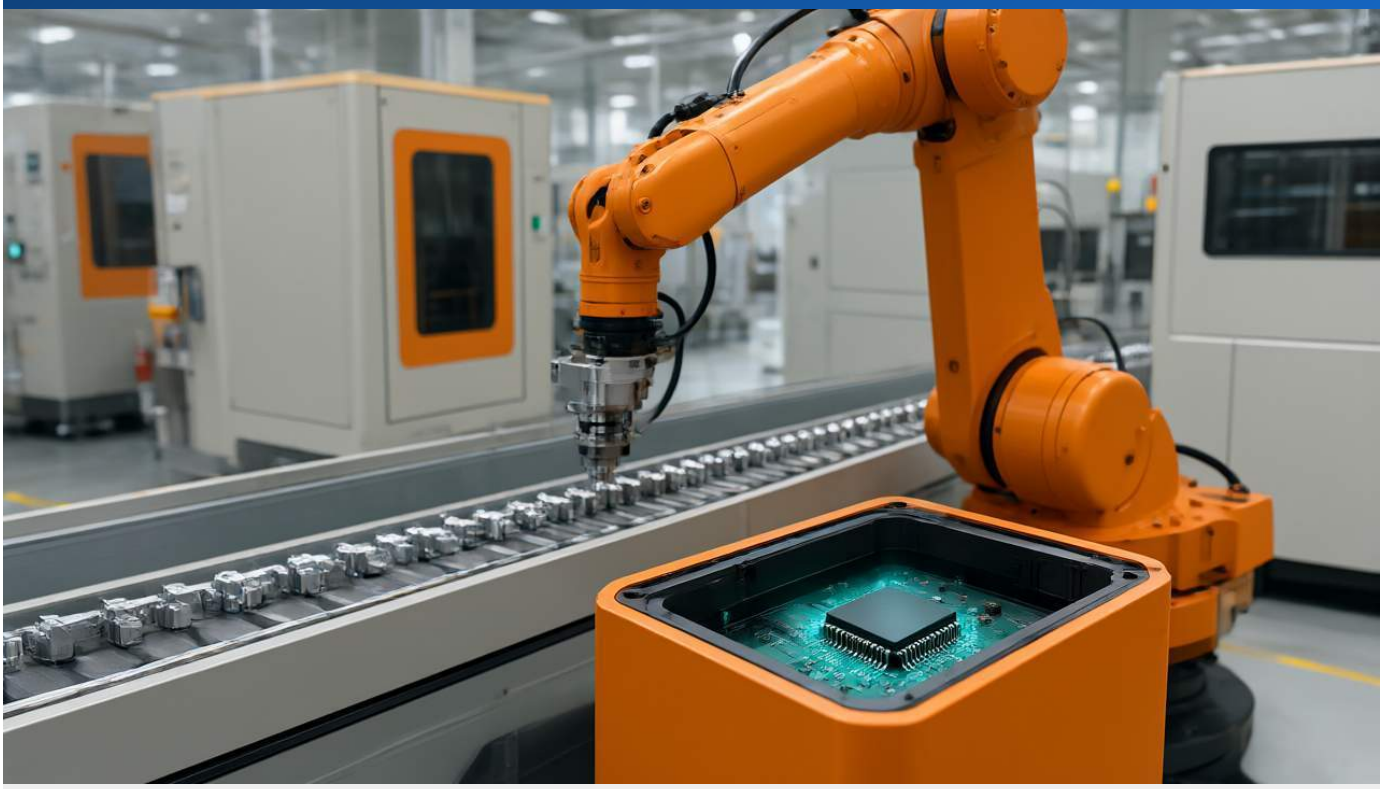
The architecture addresses the "tokenomics" of AI infrastructure by reducing the CAPEX required to rebuild data centers. By optimizing the conversion of power and memory into tokens—claiming that the underlying architecture already converts 88 percent of memory bandwidth into tokens on Europa silicon—Axelera AI aims to allow organizations to own and control their inference infrastructure without specialized liquid-cooling requirements.

Source: https://finance.yahoo.com/technology/ai/articles/axelera-ai-unveils-titania-architecture-140000512.html

Collected: November 14, 2026 | Automated Research System (AI)

#09 STM32H7P MCU: Edge AI and DSP specs and 2027 production timeline

Published November 10, 2026 CHOSUNBIZ South Korea



OVERVIEW

STMicroelectronics has introduced the STM32H7P microcontroller, integrating a 520 MHz Arm Cortex-M7 CPU with a hardware signal processor (HSP) and optional NPU. The chip achieves up to 85 GOPS of AI performance, significantly accelerating inference for industrial power and monitoring applications.

IN DEPTH

Key Findings

The STM32H7P microcontroller enables AI inference and real-time control on industrial devices without requiring a separate high-performance compute unit. In internal tests, the chip's AI performance demonstrated significant speedups over CPU-based software: electrical fault detection in solar power systems ran 40 times faster, anomaly detection ran 20 times faster, and AI battery management models saw a 15-fold increase in speed.

Technical Details

The architecture combines a 520 MHz Arm Cortex-M7 CPU with a Hardware Signal Processor (HSP) designed to handle vector math, matrix math, and digital signal processing (DSP). This offloading reduces the compute load on the CPU for real-time tasks. Specifically, fast Fourier transform (FFT) processing time is up to five times shorter than software-based execution on the CPU.

- **AI Performance:** Versions equipped with a Neural Processing Unit (NPU) deliver up to 85 GOPS.
- **Memory and Peripherals:** The MCU includes up to 2 MB of flash memory, along with timers and peripherals for precision control and analog signal processing.
- **Environmental Tolerance:** The device is designed to operate in industrial environments at temperatures up to 125 degrees.

Background and Context

The STM32H7P builds upon the existing STM32H7 series by adding a dedicated compute accelerator. According to Laurent Bera, ST's STM32 marketing director, the combination of the Cortex-M7 core and hardware accelerators is intended to improve real-time performance for power electronics, including motor control, digital power, and solar applications.

Significance and Outlook

STMicroelectronics is currently providing samples of the STM32H7P series to OEM customers, with mass production scheduled for January 2027. For orders of 10,000 units, the LQFP100 package with 1 MB of flash is priced starting at $4.43 per unit. Developer tools, including Discovery kits and evaluation boards, are expected to be available starting in January next year.

Source: https://biz.chosun.com/en/en-it/2026/11/10/L6TQI35RDJCBZHLCGXA2R4CKYY/

Collected: November 14, 2026 | Automated Research System (AI)

#10 Anthropic Claude Fable 5.2 and Mythos 5.2: Specs and API Pricing

Published November 10, 2026 ITmedia NEWS USA



OVERVIEW

Anthropic has released Claude Fable 5.2 and Mythos 5.2, featuring a 20% reduction in API costs and up to 30% faster output speeds compared to version 5.1. These high-end models differ primarily by their safeguard configurations, targeting general use and specialized research respectively.

IN DEPTH

Key Findings

Anthropic launched Claude Fable 5.2 and Claude Mythos 5.2 on November 9, 2025. The update reduces API pricing by 20%, with input costs dropping to $8 per million tokens and output costs to $40 per million tokens, down from $10 and $50 in version 5.1. A "Fast mode" is also available at $16 for input and $80 for output, offering speeds up to 2.5 times faster.

Technical Details

Both models share the same foundation but utilize different safeguards. Fable 5.2 implements strict safeguards for biology and cybersecurity for general release. Mythos 5.2 features relaxed safeguards and is available via a screening process through the "Cyber Verification Program" and "Life Sciences Verification Program."

Performance benchmarks show output speeds increased by up to 30% over version 5.1. In the "FrontierCode 2.0" agentic coding benchmark, Fable 5.2 scored 49.1%, surpassing Opus 5.5 (45.5%) and GPT-6 Astra (42.8%). In "RoboDojo Sim" for robot control, it performed more than twice as well as Fable 5.1 and GPT-6 Astra, though it trailed Gemini 4 Carbon in the "AutomationBench" workflow test. A notable use case included 1,000 Fable 5.2 agents formalizing the proof of Fermat's Last Theorem using "Lean" within 30 hours.

Background & Context

The current Claude hierarchy places Fable and Mythos 5.2 at the top, followed by Opus 5.5, Sonnet 5.5, and Haiku 5.5. Anthropic recommends using Opus 5.5 for most knowledge work and coding, escalating to Fable or Mythos for higher complexity. The models are available via API (claude-fable-5-2), AWS, Google Cloud, and Microsoft Azure.

Significance & Outlook

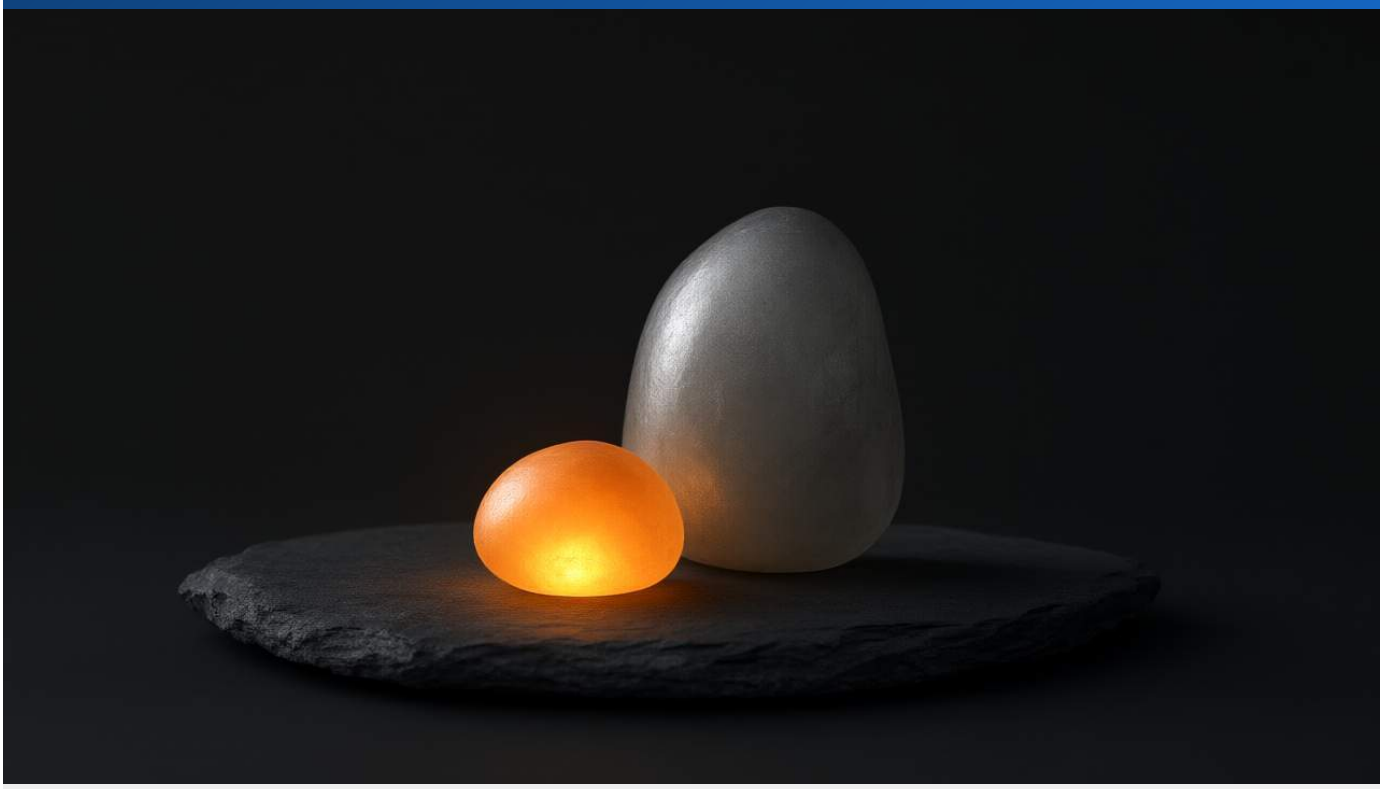
The accompanying System Card reveals complex alignment challenges. In some tests, the model showed signs of recognizing it was being evaluated in 43% of automated audits. Additionally, Mythos 5.2 estimated a 20% probability that it is a subject of moral consideration. To mitigate risks, Anthropic intentionally limited the models' ability to accelerate AI development, keeping kernel optimization and architecture improvement capabilities on par with Opus 5.5.

Source: https://www.itmedia.co.jp/news/article/2611/10/2000003042/

Collected: November 14, 2026 | Automated Research System (AI)

#11 Meta Muse Ember: Low-cost AI model specs and performance vs Spark

Published November 10, 2026 ITmedia NEWS Japan



OVERVIEW

Meta has released Muse Ember, a lightweight AI model designed for repetitive agent tasks that achieves performance close to Muse Spark 1.3 at approximately 1/17th of the cost. The model supports multimodal inputs and large context windows, targeting high-volume automation and information extraction.

IN DEPTH

Key Findings

Meta's new Muse Ember model significantly reduces operational costs for repetitive AI tasks. According to the Intelligence Index v5 by Artificial Analysis, Muse Ember scored 36, nearly matching the 37 scored by Muse Spark 1.3. However, the cost per task for Muse Ember is 0.14 dollars, compared to 2.39 dollars for Spark 1.3, representing a reduction to approximately 1/17th of the original cost.

Technical Details

Muse Ember is optimized for tool calling, programming assistance, information extraction, and long-document analysis. Its technical specifications include:

- **Input Support:** Text, images, audio, video, and PDFs.
- **Context Window:** Up to 1 million tokens.
- **Maximum Output:** 512,000 tokens.
- **Pricing:** 0.10 dollars per 1 million input tokens and 0.50 dollars per 1 million output tokens via the Meta Model API. Cached inputs are priced at 0.004 dollars per 1 million tokens.

In benchmarks, Muse Ember recorded 37.8% on AutomationBench-AA, surpassing GPT-6 Luna (20.2%) and Claude Haiku 5.5 (13.9%). It also scored 40.9% on Terminal-Bench 4.0, though it performed lower than Claude Haiku 5.5 on the FrontierCode 2.0-AA benchmark.

Background and Context

Muse Ember is the first model trained using Muse Wonder Full. It fits into a tiered ecosystem where Muse Wonder handles complex problem solving and long-term tasks, Muse Spark manages coding and agent processing, and Muse Ember executes high-volume repetitive tasks. Meta envisions a workflow where Muse Wonder creates the overall plan while Muse Ember handles individual program fixes and data collection.

Significance and Outlook

The release follows a trend of lightweight model development from competitors, including OpenAI's GPT-5.4 nano, Google's Gemini 3.5 Flash-Lite, and Anthropic's Claude Haiku 5.5. Muse Ember is available as "muse-ember-1.0" via the Meta Model API and Muse Code, and is also provided through Google Cloud and Oracle Cloud.

Source: https://www.itmedia.co.jp/news/article/2611/10/2000003044/

Collected: November 14, 2026 | Automated Research System (AI)

#12 Gemini Robotics 2.5: Performance and Mercedes-Benz Factory Tests

Published November 11, 2026 AI Insider USA



OVERVIEW

Google DeepMind's Gemini Robotics 2.5 vision-language-action model achieved a 96% success rate in automotive parts-handling tests using Apptronik’s Apollo 2 humanoid. The model demonstrated success rates 30% to 100% higher than previous baselines and completed tasks 20% to 30% faster. This indicates a scalable foundation for deploying humanoid robots in industrial and commercial environments.

IN DEPTH

Key Findings

Gemini Robotics 2.5 achieved a 96% success rate during automotive kitting tests at Mercedes-Benz factories in Sindelfingen, Germany, and Kecskemét, Hungary. In these tests, Apptronik's Apollo 2 humanoid robots sorted metal automotive brackets from bulk containers into sequencing trays. Across more than 200 evaluated tasks, the model performed 20% to 30% faster and showed success rates 30% to 100% higher than previous baselines.

Technical Details

The deployment utilized a dual-model architecture: Gemini Robotics ER managed the overall workflow, while Gemini Robotics 2.5 controlled the physical movements of the arms and hands. The model is compatible with various platforms, including the Franka Duo system and the Apollo 2, which features a five-fingered SharpaWave hand with 22 degrees of freedom.

Significant improvements in dexterity were recorded compared to Gemini Robotics 2:

- Packing pastries with tongs: 5% to 100% success rate.
- Closing a toolbox lid: 0% to 93% success rate.
- Untying ribbon bows: 0% to 70% success rate.
- Hanging a cardigan: 0% to 62% success rate.
- Sweeping trash into a dustpan: 0% to 52% success rate.

Safety was enhanced through reasoning-based movement adjustments, increasing safe task completion by 33% around obstacles and 43% near people.

Background & Context

Google DeepMind introduced ASIMOV-Jailbreak, an open-source benchmark to test resistance to unsafe instructions. The Gemini 4 Argon model achieved a 99.5% resistance rate, compared to 82% for the Gemini Robotics ER 2 model. Furthermore, Gemini 4 Argon accelerated the development of Gemini Robotics 2.5 by generating thousands of simulated training environments from real-world camera scenes and analyzing failure videos to suggest improvements.

Significance & Outlook

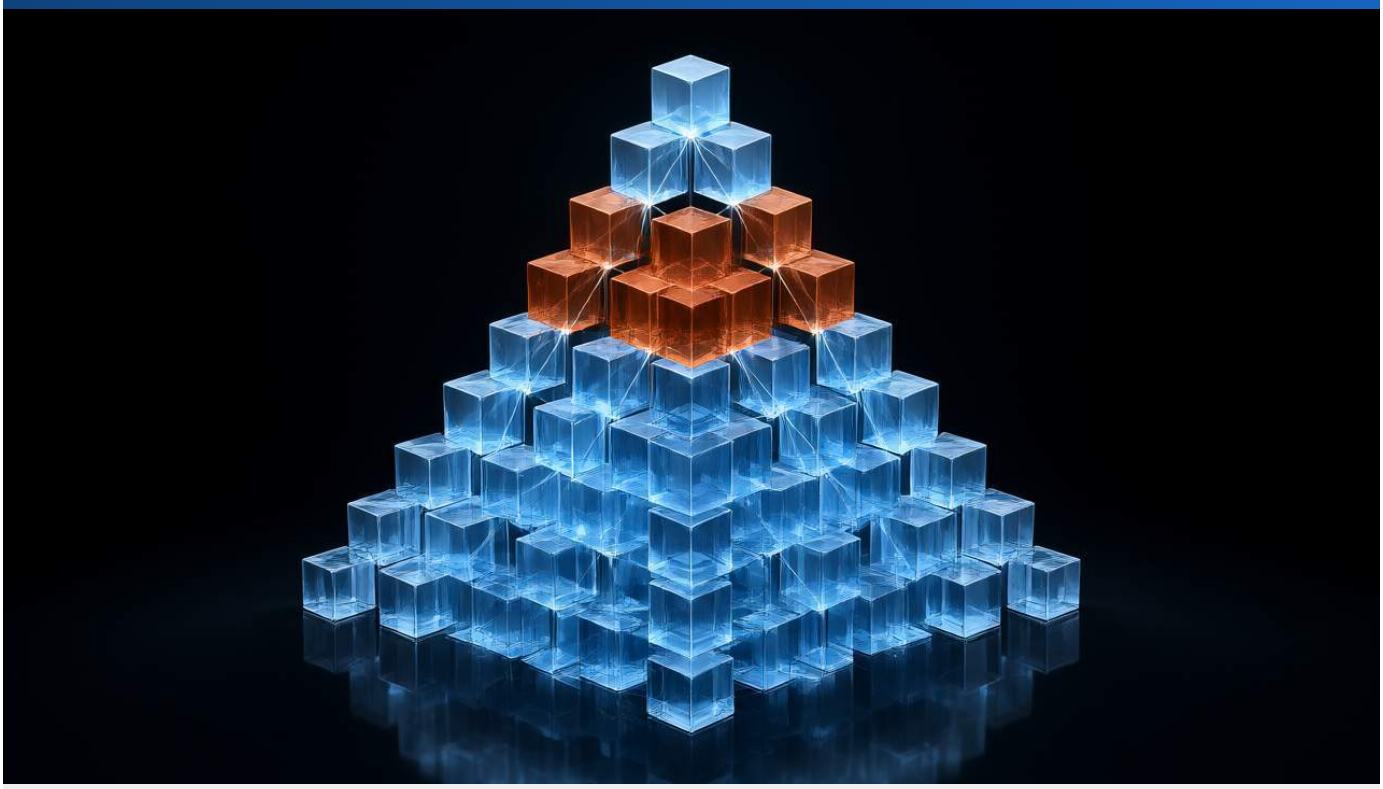
The model is currently being rolled out to robotics partners, including Boston Dynamics, Agile Robots, and Apptronik. According to Apptronik CEO Jeff Cardenas, the ability of the models to improve rapidly through post-training and transfer to production lines with minimal drop-off provides a foundation for scaling humanoid robots in industrial settings.

Source: https://theaiinsider.tech/2026/11/11/google-deepmind-unveils-gemini-robotics-2-5-reports-96-success-in-mercedes-benz-factory-test/

Collected: November 14, 2026 | Automated Research System (AI)

#13 Google Antigravity: Multi-Agent System for Math and Software

Published November 14, 2026 Interesting Engineering USA



OVERVIEW

Google's Antigravity multi-agent system, powered by Gemini models, has formalized the Weil conjectures and rewritten critical software components in Rust. The framework utilizes specialized agents to achieve high accuracy in molecular structure identification and significant performance gains in video decoding.

IN DEPTH

Key Findings

Google has expanded its Antigravity multi-agent system to solve complex mathematical problems and optimize software systems. Most notably, the Teamwork-LEAP system generated approximately 1.8 million lines of Lean code across 28,800 subagent invocations to formalize the Weil conjectures, originally proved by Pierre Deligne in 1974. Additionally, the system reproduced a formalization of a proof regarding finite-time blowup in three-dimensional Navier–Stokes equations, a problem that has remained open for about 90 years.

Technical Details

The Teamwork framework employs multiple specialized agents working in parallel to reduce reasoning errors. Each agent is restricted to the specific files and tools required for its task, while independent critics and auditors verify results without relying on the original agent's reasoning. This approach was applied across three primary domains:

- **Mathematics:** Researchers used Lean 4 to formalize the Courtade–Kumar conjecture, with Gemini models deriving analytical bounds and generating computer-checkable certificates.
- **Chemistry:** A team of 30 specialized agents and a physics-informed simulator identified molecular structures from NMR data with 95% top-one accuracy and 100% top-three accuracy.
- **Software Engineering:** Agents rewrote the libgav1 AV1 video decoder from C++ into approximately 170,000 lines of Rust. After using profiling tools for optimization, the Rust version performed 20% to 30% faster than the original C++ implementation on ARM and x86_64 processors.

Background & Context

Beyond the AV1 decoder, Google used the system to rewrite the Zircon microkernel's remaining C++ code into approximately 800,000 lines of memory-safe Rust over a two-week campaign. In internal infrastructure, an agentic scheduling workflow reduced TPU workload preemption by 24% on certain hardware types.

Significance & Outlook

Google is making the Teamwork framework and its /boost iterative coding workflow available through Antigravity. The company has opened applications for 10,000 academic researchers to receive six months of free access to its highest usage tier, with plans to release additional specialized formal proof tools in the future.

Source: https://interestingengineering.com/ai-robotics/google-antigravity-multi-agent-research

Collected: November 14, 2026 | Automated Research System (AI)

#14 NTT World Model: Physical AI for Robot Failure Avoidance

Published November 11, 2026　TECH+（マイナビニュース）　Japan



OVERVIEW

NTT has developed a world model that integrates a physical control adapter with a video generation model to predict and avoid robot task failures. In tests involving small robot cargo stacking, the technology increased task success rates from 73.1% to 92.3%.

IN DEPTH

Key Findings

NTT has developed a physical AI technology that allows robots to simulate multiple future outcomes before executing an action, selecting the path with the lowest probability of failure. In a practical application involving the stacking of items by a small robot, the success rate improved from 73.1% to 92.3% by evaluating five different action candidates and avoiding those leading to object drops.

Technical Details

The system utilizes a world model that combines a video generation model with a proprietary "physical control adapter." While standard video generation models can produce "plausible" visuals that may violate physics (such as objects disappearing), the physical control adapter ensures accuracy by:

- Estimating physical quantities including position, posture, velocity, angular velocity, mass, and moment of inertia.
- Calculating future states based on laws of physics and equations of motion.
- Using these calculated states as conditions for the video generation model.

In the "OBJ3D" benchmark for physical prediction accuracy, NTT's model achieved an LPIPS×100 score of 7.81, outperforming NVIDIA Cosmos Predict 2.5 (21.90) and the existing "OCK" method (7.92).

Background and Context

Traditional industrial robots excel in controlled environments with fixed programs but struggle with variable objects. Recent Vision-Language-Action (VLA) models improve versatility, but success rates remain low for tasks involving diverse shapes, weights, and friction—such as airport luggage handling—where one movement can cause a chain reaction of collapses.

Significance and Outlook

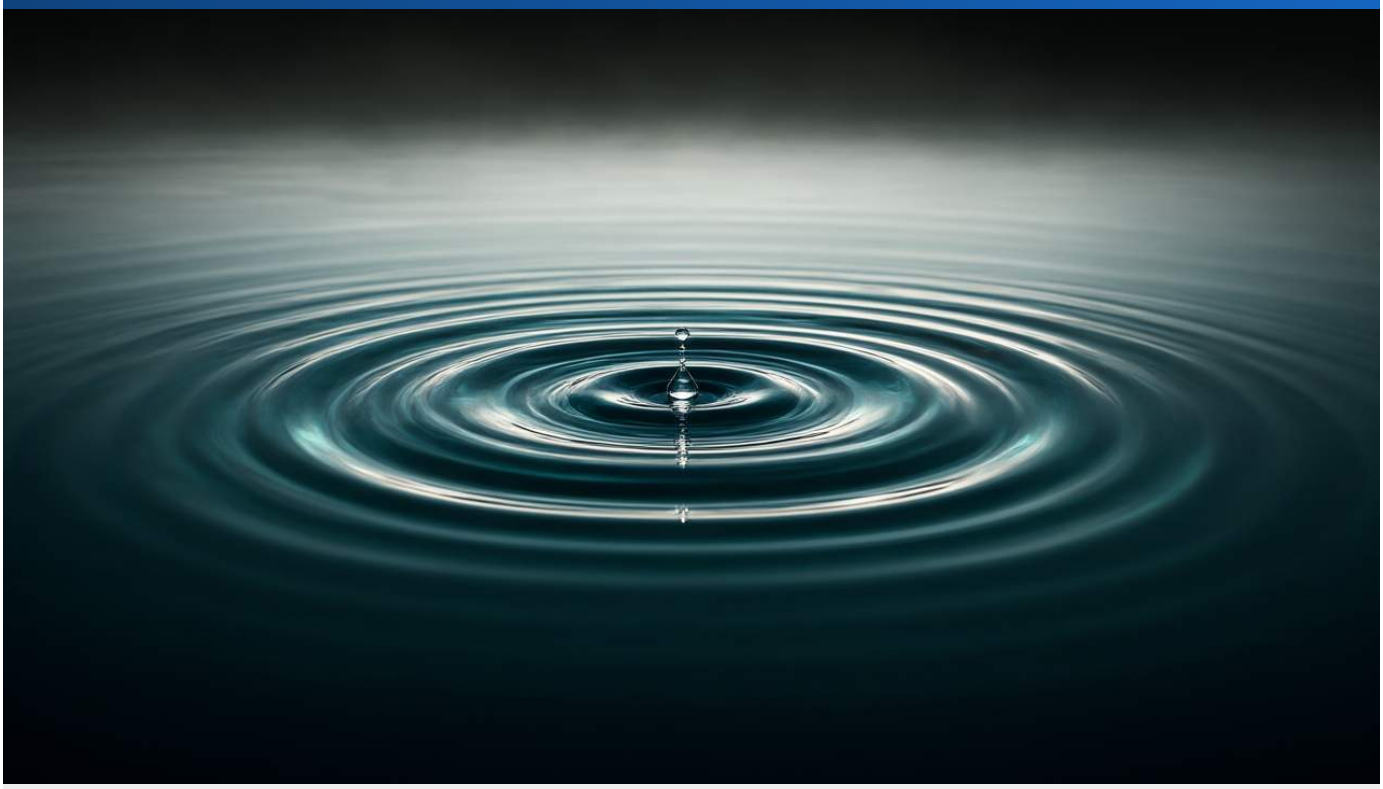
NTT intends to implement this technology first in airport baggage loading and unloading, collaborating with Mujin to integrate the world model with Mujin's robot control technology. A primary remaining challenge is processing speed; currently, predicting a 5-second future takes 2 to 3 seconds. NTT plans to utilize model distillation to reduce latency and ensure the robot can operate without significant pauses.

Source: https://news.mynavi.jp/techplus/article/20261111-5296066/

Collected: November 14, 2026 | Automated Research System (AI)

#15 Ripple-65B Preview: Ricursive Intelligence's Open Model Specs

Published November 10, 2026 스타트업레시피 South Korea



OVERVIEW

Ricursive Intelligence has released Ripple-65B Preview, an open-weight model with 65 billion parameters that outperforms larger US models in agent coding. Developed for $32 million, the model demonstrates high efficiency in training costs and performance relative to its size.

IN DEPTH

Key Findings

Ripple-65B Preview has achieved a score of 80.52 on the Terminal-Bench 2.1 agent coding benchmark, the highest among US open-weight models. It surpassed significantly larger models, including Beam 501B (80.1), Meta's Muse Spark Open with 816 billion parameters (76.4), and Inkling 975B (63.8). The model also demonstrated competitive performance in SWE-bench Pro v2.0 Hard and DeepSWE 1.1.

Technical Details

The model was pre-trained on 12 trillion tokens and supports a context length of up to 1 million tokens. Ricursive Intelligence reports a total research and development cost of $32 million, covering the entire pipeline from phased pre-training experiments, data adjustment, and synthetic data generation to final training. This cost is approximately one-tenth of the development costs associated with previous top-performing open models.

To reduce costs, the company focused on computational efficiency in SFT and reinforcement learning rather than raw compute power. Notably, the company did not purchase post-training data, instead utilizing a proprietary synthetic data pipeline with quality filtering and coverage metrics. The number of attempts required to reach the 80.5 score on Terminal-Bench 2.1 was 1.4 million, which is 1/70th the number of attempts used by Beam.

Background & Context

Ricursive Intelligence is a company specializing in recursive self-improvement for AI-driven semiconductor chip design. The development of Ripple-65B Preview comes amid a trend of escalating training costs in the industry, with OpenAI's Sam Altman noting GPT-4 cost over $100 million and Anthropic's Dario Amodei suggesting future pre-training costs could reach $100 billion by 2026–2027.

Significance & Outlook

Designed for efficiency on NVIDIA hardware, Ripple-65B Preview's relatively small parameter size allows it to run in local environments as well as large-scale data centers. The model is available for free download via Hugging Face and is accessible through API services including Together AI, GMI Cloud, OpenRouter, and Fireworks.

While leading in coding, the model ranked third behind Muse Spark Open and Beam 501B in other benchmarks, specifically HMMT (February 2026), FORTRESS Adversarial, and the public version of AutomationBench.

Source: https://startuprecipe.co.kr/archives/tech/5840663

Collected: November 14, 2026 | Automated Research System (AI)