AI News Digest

11/02/2026 - 11/08/2026

WEEKLY DIGEST: BIOLOGICAL & BIOMEDICAL SCIENCES

AI Agents Tackle End-to-End Biomedical Challenges

A wave of new benchmarks and tools is charting the capabilities of AI agents across the full spectrum of biomedical research and clinical practice. From RareDiseaseBench's 35 complex rare-disease cases to BioBasicsBench's 100 wet-lab tasks, AI systems are now required to request, interpret, and integrate heterogeneous data types. Parallel efforts such as BioSecBench-RespiratoryVirus and TxBench-Antibody evaluate agents on viral surveillance and antibody-development decisions, while foundational models like CheMeleon and Google Antigravity's cheminformatics suite streamline molecular prediction and workflow automation. Together, these advances signal a shift toward AI-driven, end-to-end decision making in biology and medicine.

KEY FINDINGS

✓ Four new benchmark suites (RareDiseaseBench, BioBasicsBench, BioSecBench-RespiratoryVirus, TxBench-Antibody) quantify AI agents' ability to handle heterogeneous biomedical data and decision points.

✓ AI agents demonstrated measurable performance across 21 models and 34 model-harness configurations on tasks ranging from rare-disease diagnosis to respiratory virus surveillance.

✓ Open-source models such as CheMeleon (trained on 1 million molecules and 1,600 descriptors) and the Google Antigravity Cheminformatics Skills suite enable accurate property prediction and automated workflow orchestration.

IMPLICATIONS

→ Standardized, task-specific benchmarks will accelerate the creation of trustworthy, regulatory-ready AI agents for clinical and research pipelines.

→ Proven ability to integrate wet-lab evidence suggests AI can shorten experimental cycles, improving speed and reproducibility of discovery.

→ Successes in AI-guided antibiotic and antibody development point to a future where AI partners are integral to therapeutic innovation, prompting new ethical and oversight frameworks.

> Key Metrics: 35 clinically challenging rare-disease cases in RareDiseaseBench; 100 wet-lab tasks in BioBasicsBench; 110 AI-driven respiratory-virus questions in BioSecBench-RespiratoryVirus; 21 models evaluated across 34 model-harness configurations; CheMeleon pre-trained on 1 million molecules and 1,600 chemical descriptors; 60 real-world antibody-development decisions in TxBench-Antibody; 34 model-harness configurations tested on TxBench-Antibody

1. RareDiseaseBench: Can AI Agents Diagnose Rare Disease?

Category: Research
Departments: Computer Science, Biological Sciences, Public Health Sciences
Topics: Healthcare AI, Multimodal Agent, Rare Disease Benchmark, Genomic Data Integration, Token Efficiency

Impact Score: 27.0
Rubric Scores

Research Relevance: 5	Educational Value: 4	Innovation / Novelty: 5
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 4

The authors introduced RareDiseaseBench, a benchmark of 35 clinically challenging rare-disease cases that requires AI agents to request and integrate heterogeneous data (genome sequencing, RNA-seq, imaging, etc.) before producing a diagnosis and management plan. Across 31 model-harness configurations and 3,255 trial runs, Gemini 4 Argon on the Pi harness achieved the highest preliminary pass rate of 43.8 %, outperforming Opus 5.5 / Claude Code and GPT-6 Astra / Pi, which reached 33.3 % (the latter at half the compute cost). Performance correlated with model strength and token efficiency, while weaker models frequently failed to request appropriate data or to process large sequencing files efficiently. In total, 29 of the 35 tasks were solved by at least one configuration, with 16 tasks passed in all three attempts by some model.

https://blog.latch.bio/p/rarediseasebench-can-ai-agents-diagnose

2. Regina Barzilay and Jim Collins win Pritzker Prize for AI and Science Research Excellence for work in AI-enabled antibiotic discovery

Category: Research
Departments: Computer Science, Pharmaceutical Sciences, Biological Sciences
Topics: Healthcare AI, Deep Learning Antibiotic Discovery, Generative Molecule Design, Narrow-Spectrum Antibiotic Prediction

Impact Score: 26.0
Rubric Scores

Research Relevance: 4	Educational Value: 4	Innovation / Novelty: 5
Practical Impact: 5	Interdisciplinary Potential: 5	Ethical / Policy Implications: 3

Regina Barzilay and James Collins received the 2026 Pritzker Prize for AI in Science Research Excellence for their joint AI-enabled antibiotic discovery work. Their 2020 Cell paper introduced halicin, the first antibiotic discovered by deep learning, identified from a model trained on 2,500 molecules and shown to kill multiple drug-resistant bacteria within weeks. In 2023 they applied the same approach to discover abaucin, a narrow-spectrum antibiotic effective against Acinetobacter baumannii, reducing resistance risk and preserving gut microbiota. The award highlights the shift from trial-and-error to generative AI design in antibiotic research, though the researchers note AI currently automates parts of the pipeline rather than fully transforming drug discovery.

https://news.mit.edu/2026/regina-barzilay-jim-collins-win-pritzker-prize-ai-and-science-research-excellence-1102

3. Measuring how AI agents handle everyday wet-lab science

Category: Research
Departments: Computer Science, Biological Sciences
Topics: Science & Research, Wet-lab Image Interpretation, Protocol Reasoning, Evidence Sufficiency Detection, Multimodal Lab Benchmark

Impact Score: 26.0
Rubric Scores

Research Relevance: 5	Educational Value: 4	Innovation / Novelty: 5
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 3

BioBasicsBench is a new benchmark, co-developed with biotech firm Ligo, that evaluates AI agents on 100 wet-lab tasks requiring interpretation of raw experimental evidence such as gel images, colony plates, microscopy and instrument readouts. The tasks, created from scientists' own data and reviewed by internal and external experts, test error detection, data interpretation, protocol understanding and the ability to recognize insufficient evidence. At the "xhigh" reasoning setting, GPT-6 Astra achieved 66.7% of available credit, outperforming GPT-5.6 Sol's 43.8%, indicating progress but also highlighting remaining challenges in image analysis, control checking and contextual reasoning.

https://openai.com/index/measuring-how-ai-agents-handle-wet-lab-science

4. BioSecBench-RespiratoryVirus-Surveillance: Can AI Agents Track Respiratory Viruses?

Category: Research
Departments: Computer Science, Public Health Sciences, Biological Sciences
Topics: Healthcare AI, AI Virus Surveillance, Pathogen Discovery LLMs, Population Surveillance Benchmark

Impact Score: 26.0
Rubric Scores

Research Relevance: 5	Educational Value: 4	Innovation / Novelty: 4
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 4

The authors introduced BioSecBench-RespiratoryVirus-Surveillance, a benchmark of 110 AI-driven questions covering diagnostic assays, pathogen discovery, viral evolution, transmission, and population surveillance. Testing 21 models across 34 model-harness configurations in 11,220 trials yielded a maximum overall pass rate of 25.5% (Opus 5.5 on Pi), with only 74 of the 110 questions solved by any setup. Population-surveillance tasks were the most challenging, achieving an average pass rate of 4.3% versus 16.7% for pathogen-discovery tasks. Model cost did not predict accuracy, as low-cost runs (e.g., Nemotron 3 Ultra at $0.22 per run) performed comparably to more expensive alternatives.

https://blog.latch.bio/p/biosecbench-respiratoryvirus-surveillance

5. How to predict the behavior of molecules from scant data

Category: Research
Departments: Computer Science, Chemistry & Biochemistry, Pharmaceutical Sciences
Topics: Science & Research, Pretrained Molecular Model, Few-Shot Property Prediction, ChemProp Fine-Tuning, OpenADMET Benchmark

Impact Score: 24.0
Rubric Scores

Research Relevance: 4	Educational Value: 4	Innovation / Novelty: 4
Practical Impact: 5	Interdisciplinary Potential: 5	Ethical / Policy Implications: 2

A MIT team introduced CheMeleon, an open-source machine-learning model that pre-trains on one million molecules and 1,600 chemical descriptors to predict diverse properties, including ligand affinities, with limited target-specific data. The method outperformed existing approaches in selecting promising compounds, as demonstrated in a September 14 paper in the Journal of Chemical Information and Modeling. CheMeleon's performance led to its frequent use by pharmaceutical companies and its inclusion in nine of the top-20 entries in the OpenADMET property-prediction contest. The model is distributed through the ChemProp package, allowing users to fine-tune it for their own property-prediction tasks.

https://news.mit.edu/2026/how-to-predict-molecule-behavior-with-scant-data-1106

6. TxBench-Antibody Development: Can AI Agents Get an Antibody to Its First Human Dose?

Category: Research
Departments: Computer Science, Pharmaceutical Sciences, Biological Sciences
Topics: Healthcare AI, AI Agent Benchmark, Antibody Development Modeling, Stage-wise Decision Optimization, Cost-Accuracy Tradeoff

Impact Score: 22.0
Rubric Scores

Research Relevance: 4	Educational Value: 3	Innovation / Novelty: 4
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 2

The authors introduced TxBench-Antibody Development, a benchmark comprising 60 real-world antibody-development decisions spanning seven stages, each paired with the underlying records and a fixed answer key. They evaluated 34 model-harness configurations across 6,110 attempts, finding the highest overall pass rate of 39.4 % for Grok 4.7 on Grok Build, with performance dropping to 13 % on toxicology and first-dose decisions. Although 57 of the 60 decisions were correctly solved by at least one setup, no model achieved high accuracy across all stages, and cost per attempt did not correlate with accuracy.

https://blog.latch.bio/p/txbenchantibody-development-can-ai

7. Agentic cheminformatics with Google Antigravity

Category: Research
Departments: Computer Science, Chemistry & Biochemistry, Pharmaceutical Sciences
Topics: Healthcare AI, Cheminformatics Automation, Multimodal Molecular Representation, Molecular Rigidification, Deterministic Tool Integration

Impact Score: 17.0
Rubric Scores

Research Relevance: 2	Educational Value: 2	Innovation / Novelty: 3
Practical Impact: 4	Interdisciplinary Potential: 5	Ethical / Policy Implications: 1

The authors released a suite of Cheminformatics Skills for Google Antigravity that integrates RDKit, ChEMBL, and PubChem to automate workflow orchestration, database navigation, and scientific reasoning while delegating molecular calculations to deterministic tools. The system unifies diverse molecular representations-text, images, and data files-enabling collaborative use by computational and experimental chemists without manual data wrangling. A technical walkthrough demonstrates the suite's application to the medicinal-chemistry strategy of rigidifying drug molecules, illustrating how cyclization reduces rotatable bonds and improves binding affinity. The case study contrasts the first-generation inhibitor crizotinib with its rigidified successor lorlatinib, showing the potency gain from eliminating molecular flexibility.

https://deepmind.google/blog/agentic-cheminformatics-with-google-antigravity/