AI Link Collection Report - 2025-07-08

Generated on July 08, 2025 at 04:24 PM

Total Items: 6

Report Statistics	
Total Links Processed	6
Generation Date	2025-07-08
Generation Time	16:24:04

Table of Contents

■ Link Collection Details

1. The Era of Exploration

■ https://yidingjiang.github.io/blog/post/exploration/

■ Full Article Content:

The Era of Exploration

- **Large Language Models (LLMs)**

- Result from three decades of freely accessible human text online.

- Ilya Sutskever compares this information reservoir to **fossil fuel**: abundant but finite.

- Studies suggest that frontier labs may exhaust high-quality English web text before the decade ends.

- Current models consume data faster than humans can produce it.

- **Era of Experience**

- Coined by David Silver and Richard Sutton.

- Progress depends on data generated by learning agents themselves.

- The bottleneck is not just any experience, but the **right kind of experience** that benefits learning.

- Future AI progress will focus on **exploration**: acquiring new and informative experiences.

- **Cost of Experience Collection**

- Scaling involves resources such as:

- Compute cycles

- Synthetic-data generation

- Data curation pipelines

- Human oversight

- Any expenditure that creates learning signals.

- All costs are simplified into a unit called **flops**:

- A flop is one floating-point operation.

- Used as a common currency to measure effort consumed by systems.

- Discussion focuses on relative spend, not specific resources.

- **Exploration in Data-Driven Systems**

- Exploration is crucial for every data-driven system to decide which experiences to collect.

- Broader definition of exploration beyond reinforcement learning (RL).

- Inspired by Minqi's article: **General intelligence requires rethinking exploration**.

- **Post Organization**

- The following sections will cover:

1. How pre-training solved part of the exploration problem.

2. Why better exploration leads to better generalization.

3. Where to allocate the next hundred thousand GPU-years.

- **Pretraining as Exploration**

- Standard LLM pipeline:

- Pretrain a large model on next-token prediction using extensive text.

- Fine-tune the model with RL for desired objectives.

- Without pretraining, RL struggles to progress.

- Smaller models can improve reasoning through distillation from larger models, suggesting:

- Large scale is not a prerequisite for effective reasoning.

- However, the need for distillation indicates that model capacity is not the only factor.

- **Exploration Tax**:

- Pretraining incurs a significant upfront cost to enhance exploration capabilities.

- Models without pretraining find it harder to explore the solution space effectively.

- Pretraining invests in diverse data to learn a rich sampling distribution for correct continuations.

- Distillation allows smaller models to inherit exploration capabilities from larger models.

The Era of Exploration

Importance of Pre-Paid Exploration

- **Smaller pretrained models** struggle to explore the solution space effectively.

- **Pretraining** involves significant compute resources to learn a rich sampling distribution for likely correct continuations.

- **Distillation** allows smaller models to inherit the exploration capabilities from larger models.

Reinforcement Learning (RL) Loop

- The general RL loop consists of:

- **Exploration**: The agent generates randomized exploration trajectories.

- **Reinforce**: Good trajectories are up-weighted; bad ones are down-weighted.

- For effective learning:

- The agent must generate a minimal number of **“good” trajectories** during exploration (known as **coverage** in RL).

- In **Large Language Models (LLMs)**, exploration is achieved through sampling from the model’s autoregressive output distribution.

- Correct solutions must be likely in the naive sampling distribution.

- Lower-capacity models may struggle to find valid solutions through random sampling, leading to ineffective reinforcement.

Challenges of Exploration

- Exploration without prior information is difficult:

- In tabular RL, every state and action can be listed, but learning requires many trials.

- A known lower-bound on sample complexity is:

- $\Omega(\frac{SAH^2}{\epsilon^2})$ (Dann & Brunskill, 2015)

- Where:

- **S** = size of the state space

- **A** = size of the action space

- **H** = horizon

- ε = distance to the best solution

- Minimum episodes grow linearly with state-action pairs and quadratically with the horizon.

- For LLMs, the state space includes all possible text prefixes, and the action space includes any next token, making exploration nearly impossible without prior information.

Role of Pretraining

- Pretraining has largely handled the exploration challenge by learning a better prior for sampling trajectories.

- However, this constrains the types of trajectories that can be sampled naively.

- To advance, we need to explore beyond the prior.

Exploration and Generalization

- Historically, RL research focused on single environments (e.g., Atari, MuJoCo):

- This is akin to training and testing on the same data point.

- Performance in a single environment does not indicate how well a model can handle novel situations.

- **Generalization** is crucial in machine learning:

- Success on unseen problems is more important than solving known issues.

- Generalization performance is critical for LLMs:

- During training, LLMs see a limited set of prompts but must handle diverse user queries at deployment.

- Current LLMs perform well on tasks with verifiable rewards (e.g., coding puzzles) but struggle with ambiguous tasks (e.g., writing a novel).

Options for Training Generalizable Models

- **Data diversity** is key for robust generalization in deep learning.

- Exploration influences data diversity:

- In supervised learning, each labeled example reveals all details in one pass, requiring more data for diversity.

- In RL, each interaction reveals a narrow slice of the environment.

- Agents must collect varied trajectories to build a representative picture.

- Lack of diversity in collected trajectories can lead to overfitting within the same environment.

The Era of Exploration

Supervised Learning vs. Reinforcement Learning (RL)

- **Supervised Learning**:

- Labeled examples reveal all details in a single forward pass.

- Data diversity can only be increased by collecting more data.

- **Reinforcement Learning (RL)**:

- Each interaction exposes a narrow slice of the environment.

- Agents must gather varied trajectories to build a representative picture.
- Lack of diversity in trajectories (e.g., naive random sampling) can lead to overfitting.

Challenges with Multiple Environments

- **Procgen Benchmark**:
- A collection of Atari-like games with procedurally generated environments.
- Each game theoretically contains “infinitely” many environments.
- Objective: Train on a fixed number of environments and generalize to unseen ones.

Current Approaches and Limitations

- Many approaches treat the problem as a **representation learning** issue:
- Apply regularization techniques from supervised learning (e.g., dropout, data augmentation).
- These techniques help but often overlook **exploration**, a crucial component of RL.

- **Exploration Strategies**:
- Agents can improve generalization by changing exploration methods.
- Previous work showed that pairing RL algorithms with stronger exploration strategies can:
- Double generalization performance on Procgen without explicit regularization.
- Allow models to leverage more expressive architectures and resources.

Exploration in LLMs

- While Procgen is simpler than current LLM challenges, the problem structure is similar:
- RL agents are trained on a finite set of problems and tested on new ones without further training.
- Current exploration methods for LLMs are basic:
- Typically involve sampling from the model’s autoregressive distribution with tweaks (e.g., temperature, entropy bonus).
- There is significant potential for better exploration approaches.

- **Challenges in Exploration**:
- Few successful examples of improved exploration strategies.
- Possible reasons for limited success:
- Difficulty of the problem.
- Inefficiency in computational resources.
- Lack of effort in exploring new methods.
- If Procgen-style exploration gains translate, we may be missing out on efficiency and new capabilities.

Two Axes of Scaling Exploration

- **Exploration**: Deciding what data the learner will see occurs on two axes:

1. **World Sampling**:
- Refers to the specific problem to be solved.
- In supervised learning, it involves data collection, synthetic generation, and curation.
- In RL, it involves designing or generating environments (e.g., math puzzles, coding problems).
- Arranging worlds into curricula is also possible.

- Determines the limit on the information any agent can learn.

2. **Path Sampling**:

- Unique to RL; involves deciding how to gather data within a chosen world.

- Agents select trajectories to collect (e.g., random walks, curiosity-driven policies, tree search, tool-use).

- Different strategies incur varying computational costs and produce different training distributions.

- Essentially, it's about what the learner "wants" to see.

- In supervised learning or unsupervised pretraining, the second axis incurs a constant cost:

- A single forward (and backward) pass accesses all information in each data point.

- Exploration cost primarily resides in the first axis (world sampling).

- Computational resources can be allocated to acquiring new worlds or processing existing ones.

The Era of Exploration

- **Supervised Learning vs. Reinforcement Learning (RL)**

- In supervised learning or unsupervised pretraining:

- The second axis incurs a **constant cost** due to access to all information in each data point (e.g., **cross-entropy loss**).

- Exploration cost primarily resides on the **first axis** – **world sampling**.

- Flops can be allocated to:

- Acquiring new worlds (new data points)

- Processing existing worlds (curation and synthetic data)

- **Flexibility in RL**

- RL offers more flexibility on both axes:

- Random trajectories often reveal little about ideal behavior.

- Information density (useful bits per flop) in RL is lower than in supervised learning or pretraining.

- Naïve trajectory sampling risks wasting flops on noise.

- Important to be judicious in spending flops:

- Options include:

- Sampling more trajectories from a single environment.

- Allocating flops to strategize the next trajectory for discovering high-value states and actions.

- **Maximizing Information per Flop**

- High-level goal in machine learning:

- Maximize information per flop.

- Two levers create a **trade-off curve**:

- Excessive resources on world sampling without path sampling may yield no meaningful experience.

- Over-investing in a small set of worlds may lead to overfitting and lack of generalizable behavior.

- Ideal scenario:

- Balanced resource allocation between sampling new worlds and running algorithms to extract more information from a single world.

- **Chinchilla Scaling Laws**

- Similar to Chinchilla scaling laws:

- Two axes correspond to compute used for different types of sampling rather than parameters and data.

- At each performance level, trace an **isoperformance curve**:

- X-axis: Compute for interacting with environments.

- Y-axis: Compute for generating or running environments (e.g., generative verifier with CoT).

- **Path Sampling vs. World Sampling**

- Path sampling is a well-defined problem:

- Objective: Reduce model uncertainty.

- Existing exploration approaches have strong sample complexity but can be expensive.

- World sampling objectives are less clear:

- Open-ended learning requires defining the universe of environments or a subjective observer to judge interesting outcomes.

- **Challenges in World Sampling**

- Objective for world sampling:

- The space of environments is **infinite**, but resources are **finite**.

- Must express preferences over environments for useful outcomes.

- Designing environments may resemble selecting pretraining data:

- Hard to determine why one environment aids another.

- Need a variety of environments.

- Likely scenario:

- Designing specs within individual expertise or domain of interest.

- Accumulating enough “human-approved” and “useful” specs may lead to learning common principles and automating the process, similar to pretraining data selection.

- Preliminary evidence suggests fewer environments may suffice for achieving generality in decision-making.

The Era of Exploration

- **Objective**: Train an agent for general exploration and decision-making in out-of-distribution environments.

- **Design Process Acceleration**:

- Utilizing existing **Large Language Models (LLMs)** can significantly speed up the design process.

- Anticipated trend: Individuals will design specifications within their own areas of expertise.

- **Learning from Specifications**:

- Once enough "human-approved" and "useful" specifications are gathered, we can:

- Identify common principles.

- Automate the design process, similar to current pretraining data selection.

- **Generalization Concerns**:

- It may be inconvenient to require as many environments as pretraining data for achieving similar decision-making generality.

- Preliminary evidence suggests that a small number of environments can suffice for training agents in out-of-distribution scenarios.

- **Scaling Challenges**:

- Scaling exploration and decision-making is less straightforward than scaling pretraining.

- Potential solutions include:

- Reliable methods for **world sampling**.

- Intelligent approaches for **path sampling**.

- Expected outcome: Isoperformance curves that bend inward towards the origin, indicating efficient resource allocation between environments and agents.

Final Thoughts

- **Exploration as a Key Focus**:

- While many tangents (e.g., better curiosity objectives, open-endedness, meta-exploration) could be explored, the main point is:

- Existing scaling paradigms are effective but will eventually reach saturation.

- The next focus should be on where to allocate additional compute resources.

- **Promising Directions**:

- Exploration, specifically world and path sampling, is a promising avenue for future research.

- Current unknowns include:

- Appropriate scaling laws.

- Effective environment generators.

- Suitable exploration objectives.

- The coming years will determine if exploration can enhance computational efficiency beyond existing paradigms.

Acknowledgements

- Special thanks to:

- Allan Zhou

- Sam Sokota

- Minqi Jiang

- Ellie Haber

- Alex Robey

- Swaminathan Gurumurthy

- Kevin Li

- Calvin Luo

- Abitha Thankaraj

- Zico Kolter

- Their feedback and discussions were invaluable in shaping this draft.

Additional Insights

- **Alternative Possibilities**:

- The RL optimization objective may not perform well with smaller models, but this is unlikely since prior successful RL applications involved small models.

- **Information Availability**:

- Models may not fully exploit available information due to computational limitations, but the information remains accessible if desired.

- **Generalization Assumption**:

- For generalization to be feasible, a "good enough" policy must exist for all environments, akin to the assumption of minimal label noise in supervised learning.

- **Performance Benchmark**:

- Current work reportedly sets a new state-of-the-art performance on the "25M easy" benchmark of **ProcGen**.

- **Random Sampling Effectiveness**:

- Random sampling works reasonably well for many problems (e.g., **Atari**), indicating more about the environments than the exploration methods.

- **Exploration Algorithms**:

- A variety of RL algorithms, such as **posterior sampling** or **information-directed sampling**, aim to reduce model uncertainty during exploration.

- However, these methods are generally too costly for implementation at the scale of LLMs, and existing approximations have not been widely adopted.

- **Shared by:** Sai
- **Shared on:** Jul 07, 2025 at 11:21 AM
- **Domain:** yidingjiang.github.io
- **Length:** 6,325 words
- **Processed:** Jul 08, 2025 at 04:23 PM

---

2. GitHub - apple/ml-diffucoder: DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation

- `https://github.com/apple/ml-diffucoder`

Website Information:

Website Description:

The GitHub repository for DiffuCoder serves as a comprehensive resource for understanding and enhancing masked diffusion models specifically tailored for code generation. This project is closely associated with the research paper titled 'DiffuCoder: Understanding and Improving Masked Diffusion Models for Code Generation,' providing essential documentation, code files, and collaborative tools for developers and researchers interested in machine learning and code synthesis.

The repository includes various resources such as setup instructions, demo scripts, and contribution guidelines, making it an invaluable platform for both academic and practical applications in the field of artificial intelligence. Users can explore the latest updates, access the codebase, and participate in the ongoing development of the project, fostering a community of innovation and knowledge sharing in the realm of code generation technologies.

- **Shared by:** Sai
- **Shared on:** Jul 02, 2025 at 09:58 AM
- **Domain:** github.com
- **Length:** 2,901 words
- **Processed:** Jul 08, 2025 at 04:22 PM

---

3. apple/DiffuCoder-7B-cpGRPO · Hugging Face

https://huggingface.co/apple/DiffuCoder-7B-cpGRPO

Website Information:

Website Description:

The Hugging Face page for DiffuCoder-7B-cpGRPO provides an in-depth overview of a refined machine learning model designed for code generation. This variant enhances the original DiffuCoder-Instruct model through reinforcement learning techniques, specifically Coupled-GRPO, which significantly boosts performance on code generation benchmarks. Users can find detailed training recipes, usage examples, and links to relevant resources, including a research paper and a GitHub repository for further exploration.

This resource is particularly valuable for developers, researchers, and data scientists interested in advanced code generation techniques and machine learning model optimization. The page serves as a comprehensive guide for understanding the capabilities of the DiffuCoder-7B-cpGRPO model, making it an essential tool for those looking to implement cutting-edge AI solutions in their projects.

- **Shared by:** Sai
- **Shared on:** Jul 02, 2025 at 09:58 AM
- **Domain:** huggingface.co
- **Length:** 574 words
- **Processed:** Jul 08, 2025 at 04:24 PM

---

4. The Chatbot is Dead. Long Live the Orchestrator

https://community.coda.io/t/the-chatbot-is-dead-long-live-the-orchestrator/56357

Full Article Content:

The Chatbot is Dead. Long Live the Orchestrator

preload-content:

The Chatbot is Dead. Long Live the Orchestrator

The Chatbot is Dead. Long Live the Orchestrator Bill_French July 1, 2025, 3:20pm While the world was distracted by talking dolls, Grammarly quietly built an AI that could act. They didn't use a better model. They used a better weapon: Coda. While the world was distracted by talking dolls, Grammarly quietly built an AI that could act. They didn't use a better model. They used a better weapon: Coda. Your Prompts Are a Prayer to an Amnesiac God

Key points:

- Your Prompts Are a Prayer to an Amnesiac God Let's be brutally honest.
- The entire AI industry is captivated by a lie.
- We've anointed "prompt engineering" as a mystical art, a high priesthood for coaxing wisdom from silicon gods.
- We are meticulously polishing the conversational skills of a machine with terminal amnesia.
- Let's be brutally honest.

The entire AI industry is captivated by a lie. We've anointed "prompt engineering" as a mystical art, a high priesthood for coaxing wisdom from silicon gods. It's a sham. We are meticulously polishing the conversational skills of a machine with terminal amnesia. We were promised an omniscient partner, an AI co-pilot. What we got was a brilliant intern with no long-term memory, an entity we must re-brief from scratch every five minutes. This isn't productivity. It's digital babysitting for a machine that can recite Shakespeare but struggles to remember your name or perform precise calculations. We were promised an omniscient partner, an AI co-pilot. What we got was a brilliant intern with no long-term memory, an entity we must re-brief from scratch every five minutes. This isn't productivity. It's digital babysitting for a machine that can recite Shakespeare but struggles to remember your name or perform precise calculations. Today's large language models are amnesiacs by design. We celebrate their ballooning context windows—a million, two million tokens—as if it were memory. It's not. It's a bigger notepad. A volatile transcript that dissolves into the ether the moment you close the tab. This architecture condemns us to a state of digital shrapnel: your project plan lives in Slack, your research is scattered across browser tabs, your decisions are buried in email, and your draft is in a doc. We ask our AI to be intelligent, but we force it to operate blindfolded, guessing the shape of our work by touching one disconnected piece at a time. Today's large language models are amnesiacs by design. We celebrate their ballooning context windows—a million, two million tokens—as if it were memory. It's not. It's a bigger notepad. A volatile transcript that dissolves into the ether the moment you close the tab. This architecture condemns us to a state of digital shrapnel: your project plan lives in Slack, your research is scattered across browser tabs, your decisions are buried in email, and your draft is in a doc. We ask our AI to be intelligent, but we force it to operate blindfolded, guessing the shape of our work by touching one disconnected piece at a time. This makes you the AI's external hard drive. You are the connective tissue. You perform the soul-crushing labor of copying, pasting, and re-explaining context, bridging the gap with every single query. This isn't a feature. It's a catastrophic, unforgivable design flaw. This makes you the AI's external hard drive. You are the connective tissue. You perform the soul-crushing labor of copying, pasting, and re-explaining context, bridging the gap with every single query. This isn't a feature. It's a catastrophic, unforgivable design flaw. Stop Describing. Start Commanding. Stop Describing. Start Commanding. For a moment, I thought the answer was contexting—the architectural discipline of curating a rich data environment for an AI agent. It was the right instinct, but the wrong verb. It was a step away from the vacant art of prompting, but it was still just talking at the machine. For a moment, I thought the answer was contexting—the architectural discipline of curating a rich data environment for an AI agent. It was the right instinct, but the wrong verb. It was a step away from the vacant art of prompting, but it was still just talking at the machine. You cannot build a skyscraper by describing it to a pile of bricks, no matter how eloquently. You need an architectural plan, a crane, and a crew. Prompting is the

description. Coda is the crane and the blueprint. It transforms your context from a static pile of information into a dynamic set of executable instructions. You cannot build a skyscraper by describing it to a pile of bricks, no matter how eloquently. You need an architectural plan, a crane, and a crew. Prompting is the description. Coda is the crane and the blueprint. It transforms your context from a static pile of information into a dynamic set of executable instructions. You cannot build a skyscraper by describing it to a pile of bricks, no matter how eloquently. You need an architectural plan, a crane, and a crew. Prompting is the description. Coda is the crane and the blueprint. It transforms your context from a static pile of information into a dynamic set of executable instructions. The true frontier isn't what an AI knows. It's what it can do. This demands a new class of software: an Agentic Orchestration Framework. A system that doesn't just talk, but commands, coordinates, and executes. It directs multiple specialized agents—AI, automation, and human—across complex, multi-step workflows with unwavering precision. The true frontier isn't what an AI knows. It's what it can do. This demands a new class of software: an Agentic Orchestration Framework. A system that doesn't just talk, but commands, coordinates, and executes. It directs multiple specialized agents—AI, automation, and human—across complex, multi-step workflows with unwavering precision. The archetype for this framework has been hiding in plain sight: Coda. The archetype for this framework has been hiding in plain sight: Forget the "all-in-one doc" marketing. That was the Trojan horse. Coda is a workflow engine for manufacturing bespoke, agentic software without writing a line of code. Its architecture is the very blueprint for orchestration:

Key points:

- Forget the "all-in-one doc" marketing.
- That was the Trojan horse.
- Coda is a workflow engine for manufacturing bespoke, agentic software without writing a line of code.
- Its architecture is the very blueprint for orchestration: • Docs as the Command Center: The unified surface where human intent and AI execution converge.
- Docs as the Command Center: The unified surface where human intent and AI execution converge.
- • Structured Tables as the Memory: A shared, persistent "brain" that provides unwavering context, rendering the amnesiac chat log obsolete.

Structured Tables as the Memory: A shared, persistent "brain" that provides unwavering context, rendering the amnesiac chat log obsolete. • Packs as the Limbs: The API-driven connectors that give agents power over the real world—to manipulate Google Calendar, create Jira tickets, or rewrite Salesforce records. Packs as the Limbs: The API-driven connectors that give agents power over the real world—to manipulate Google Calendar, create Jira tickets, or rewrite Salesforce records. • Automations as the Nervous System: The rule-based engine that triggers actions and executes entire workflows with inhuman speed and reliability. Automations as the Nervous System: The rule-based engine that triggers actions and executes entire workflows with inhuman speed and reliability. CleanShot 2025-07-01 at 09.14.43@2x1274×842 43.6 KB

CleanShot 2025-07-01 at 09.14.43@2x 1274×842 43.6 KB This was never a document app. It's a factory for building intelligent actors. This was never a document app. It's a factory for building intelligent actors. The Grammarly Gambit: An Empire in Two Moves

Key points:

- The Grammarly Gambit: An Empire in Two Moves While the market obsessed over chatbot demos and press releases, Grammarly executed a two-step strategic coup to build the world's first true agentic productivity platform.

- Anyone who saw these as unrelated acquisitions wasn’t just missing the story; they were illiterate in the language of power.

- While the market obsessed over chatbot demos and press releases, Grammarly executed a two-step strategic coup to build the world’s first true agentic productivity platform.

- Anyone who saw these as unrelated acquisitions wasn’t just missing the story; they were illiterate in the language of power.

- Move 1 (Acquire the Brain): Seize the Orchestration Framework

In late 2024, Grammarly acquired Coda.

- They didn’t buy a popular doc app.

They bought the operating system for their future AI agents. This was the foundational act of war. As undeniable proof, Coda founder Shishir Mehrotra wasn’t just given a board seat; he was installed as Grammarly’s new CEO. He isn’t running a company; he is performing a hostile takeover of its DNA, injecting Coda’s agentic framework into a platform with 40 million daily active users. Move 1 (Acquire the Brain): Seize the Orchestration Framework

In late 2024, Grammarly acquired Coda. They didn’t buy a popular doc app. They bought the operating system for their future AI agents. This was the foundational act of war. As undeniable proof, Coda founder Shishir Mehrotra wasn’t just given a board seat; he was installed as Grammarly’s new CEO. He isn’t running a company; he is performing a hostile takeover of its DNA, injecting Coda’s agentic framework into a platform with 40 million daily active users. Move 2 (Conquer the Battlefield): Seize the Critical Interface

Months later, Grammarly acquired Superhuman. This was not about adding a slick email client. Superhuman is what Mehrotra calls the “perfect staging ground for orchestrating multiple AI agents simultaneously.” Email is the chaotic nexus where work, communication, and tasks collide. Grammarly didn’t buy Superhuman for its pathetic summarization features; they bought the most valuable turf in professional life to serve as the GUI for their Coda-powered agentic backend. Imagine it: A sales agent, a support agent, and a scheduling agent collaborating within a single email draft, orchestrated by the Coda engine, pulling live context from connected Packs, executing tasks across a dozen SaaS apps. That is the gambit. Move 2 (Conquer the Battlefield): Seize the Critical Interface

Months later, Grammarly acquired Superhuman. This was not about adding a slick email client. Superhuman is what Mehrotra calls the “perfect staging ground for orchestrating multiple AI agents simultaneously.” Email is the chaotic nexus where work, communication, and tasks collide. Grammarly didn’t buy Superhuman for its pathetic summarization features; they bought the most valuable turf in professional life to serve as the GUI for their Coda-powered agentic backend. Imagine it: A sales agent, a support agent, and a scheduling agent collaborating within a single email draft, orchestrated by the Coda engine, pulling live context from connected Packs, executing tasks across a dozen SaaS apps. That is the gambit. CleanShot 2025-07-01 at 09.17.51@2x1434×992 78 KB

CleanShot 2025-07-01 at 09.17.51@2x 1434×992 78 KB

Your AI Stack is a Museum Piece

Key points:

- Your AI Stack is a Museum Piece The chasm between the dying paradigm and the emerging one is not an increment; it is a cliff.

- One is a toy, the other is a weapon.

- One talks, the other acts.

- The stack Grammarly is building does not compete with the old one.
- It renders it irrelevant.
- The chasm between the dying paradigm and the emerging one is not an increment; it is a cliff.

One is a toy, the other is a weapon. One talks, the other acts. The stack Grammarly is building does not compete with the old one. It renders it irrelevant. The Moat Isn't the Model; It's the Machine

Key points:

- The Moat Isn't the Model; It's the Machine The winners of the AI war will not be the companies with the largest language model.
- That is a commodity race to the bottom.
- They will be the ones who own the orchestration framework that makes those models act.
- Building a better chatbot today is like perfecting the horse-drawn carriage in the age of the automobile.
- The real innovation was the assembly line and the highway system.
- The only defensible moat is the machine: the framework that enables action, the integrations that give it reach, the structured data that serves as its memory, and the user workflows that become its territory.

The winners of the AI war will not be the companies with the largest language model. That is a commodity race to the bottom. They will be the ones who own the orchestration framework that makes those models act. Building a better chatbot today is like perfecting the horse-drawn carriage in the age of the automobile. The real innovation was the assembly line and the highway system. The only defensible moat is the machine: the framework that enables action, the integrations that give it reach, the structured data that serves as its memory, and the user workflows that become its territory. Stop asking if your AI is smart. Start demanding that it act. Stop celebrating prompts. Start building engines of execution. Stop asking if your AI is smart. Start demanding that it act. Stop celebrating prompts. Start building engines of execution. The future of work isn't a conversation. It's a command. The companies that understand this are building empires. The rest are polishing tombstones. The future of work isn't a conversation. It's a command. The companies that understand this are building empires. The rest are polishing tombstones. ps. A warm welcome to Rahul and the SuperHuman team.

ps. A warm welcome to Rahul and the SuperHuman team. 15 Likes Exciting News: Grammarly to acquire Superhuman Christiaan_Huizer July 1, 2025, 4:13pm thx for sharing your ideas. interesting & promising.

what do you believe would be a next tool to acquire @Bill_French?

thx for sharing your ideas. interesting & promising.

what do you believe would be a next tool to acquire @Bill_French 2 Likes Melanie_Teh July 1, 2025, 4:26pm my $$ is on reclaim or motion… something along the lines of calendar/task management

my $$ is on reclaim or motion… something along the lines of calendar/task management edit: oh, cannot be reclaim as they got bought by dropbox

edit: oh, cannot be reclaim as they got bought by dropbox 1 Like Bill_French July 1, 2025, 4:30pm I’m uncertain, but I have to believe someone at Grammarly (who is apparently really dialed in to the way I think) is watching Pieces, Flowith, and Dia. I’m uncertain, but I have to believe someone at Grammarly (who is apparently really dialed in to the way I think) is watching Pieces Flowith 5 Likes Agile_Dynamics July 2, 2025, 11:21pm Shishir keeps talking about ‘surfaces’. Email is the most frequent surface for Grammarly usage (=>Superhuman)

Documents and sheets are another major surface (=>Coda)

Key points:

- Shishir keeps talking about ‘surfaces’.
- Email is the most frequent surface for Grammarly usage (=>Superhuman)

Documents and sheets are another major surface (=>Coda) So whats the next-biggest surface where we spend our time? So whats the next-biggest surface where we spend our time? How about messaging and chats and forums? How about messaging and chats and forums? So maybe Grammarly has it’s eye on one of those? Not the major players maybe, but a startup with oodles of innovation and AI expertise? So maybe Grammarly has it’s eye on one of those? Not the major players maybe, but a startup with oodles of innovation and AI expertise? Another key criteria for these acquisitions is a shared vision.

- It was highlighted during the Coda deal.
- It’s highlighted again by Shishir and Rahul during the Superhuman deal.
- Another key criteria for these acquisitions is a shared vision.
- It was highlighted during the Coda deal.

It’s highlighted again by Shishir and Rahul during the Superhuman deal. So the next target will have a messaging surface, an AI focus, and a clear vision that matches that of Shashir, Rahul, and their teams. So the next target will have a messaging surface, an AI focus, and a clear vision that matches that of Shashir, Rahul, and their teams. I dont know the marketplace well enough to identify candidates, but perhaps someone reading this does? I dont know the marketplace well enough to identify candidates, but perhaps someone reading this does? Just a thought. Just a thought. 2 Likes Nina_Ledid July 3, 2025, 9:03am As always, I’ve greatly enjoyed reading your perspective, @Bill_French, thanks for sharing! As always, I’ve greatly enjoyed reading your perspective, @Bill_French, thanks for sharing! I do remember from past conversations your concerns about Coda Pack’s limited support for common integration patterns (eg handling incoming webhooks, running persistent background services, or responding to external events)

I do remember from past conversations your concerns about Coda Pack’s limited support for common integration patterns (eg handling incoming webhooks, running persistent background services, or responding to external events) In your post above, you envision Coda as the blueprint for a new kind of orchestration framework. In your post above, you envision Coda as the blueprint for a new kind of orchestration framework. Do you think what once seemed like a limitation (Packs’ closed and controlled architecture) is now less of a drawback? Do you think what once seemed like a limitation (Packs’ closed and controlled architecture) is now less of a drawback? Or do you even see it potentially becoming a competitive advantage as the focus shifts from open interoperability to orchestrated execution? Or do you even see it potentially becoming a competitive advantage as the focus shifts from open interoperability to orchestrated execution? Thanks,

Nina

Thanks,

- Likes Stefan_Stoyanov July 5, 2025, 8:17pm The reason I’m not yet bullish on Coda’s future is due to one key problem that LLMs still don’t solve particularly well: data collection. Specifically, I’ve been thinking a lot about the following limitations of Coda, as I understand

its vision going forward:

Key points:

• The reason I’m not yet bullish on Coda’s future is due to one key problem that LLMs still don’t solve particularly well: data collection.

• Specifically, I’ve been thinking a lot about the following limitations of Coda, as I understand its vision going forward: • Collecting data by typing into Coda’s quadrangle notes (cells) is too slow and impractical.

• Instead, data should come from voice/video meetings and recordings, mobile/desktop screen activity, linked/shared content sources, mobile phone conversations, and visual/audible/touch perception input devices.

• Pushing data from hardware devices in real time becomes even more critical than pulling it from structured database, because if data isn’t captured quickly enough, the use of the database can become unreliable and irrelevant even within a minute.

• Collecting data by typing into Coda’s quadrangle notes (cells) is too slow and impractical.

• Instead, data should come from voice/video meetings and recordings, mobile/desktop screen activity, linked/shared content sources, mobile phone conversations, and visual/audible/touch perception input devices.

Pushing data from hardware devices in real time becomes even more critical than pulling it from structured database, because if data isn’t captured quickly enough, the use of the database can become unreliable and irrelevant even within a minute. • Email is no longer a practical communication channel - it belongs to the past. As someone who has long supported email, it’s hard for me to admit this, but emails are now like the written contracts of a previous era. Today, very little needs to be formally written when people prefer adaptable, relevant content in audio/video formats. Email is no longer a practical communication channel - it belongs to the past. As someone who has long supported email, it’s hard for me to admit this, but emails are now like the written contracts of a previous era. Today, very little needs to be formally written when people prefer adaptable, relevant content in audio/video formats. • The joint venture company’s language support is essentially limited to English at the moment. In an era of LLMs—where tokenization of language is foundational to progress—this is concerning. Many LLMs still perform poorly in languages other than English. What does it mean to build a great app for the UK but a terrible one for France, simply because UX in the UK relies on quick chat-based communication, while in France it might depend on elaborate instructions—or worse, traditional filters and search? The joint venture company’s language support is essentially limited to English at the moment. In an era of LLMs—where tokenization of language is foundational to progress—this is concerning. Many LLMs still perform poorly in languages other than English. What does it mean to build a great app for the UK but a terrible one for France, simply because UX in the UK relies on quick chat-based communication, while in France it might depend on elaborate instructions—or worse, traditional filters and search? My understanding of LLMs and Coda’s vision is limited, so I’d really appreciate the community’s thoughts on these points. My understanding of LLMs and Coda’s vision is limited, so I’d really appreciate the community’s thoughts on these points. 1 Like Christiaan_Huizer July 6, 2025, 11:53am HI @Stefan_Stoyanov

Key points:

• @Stefan_Stoyanov Thanks for sharing your thoughts, those are valid concerns.

• Just so you know, I don’t have access to any internal “cookbook” or strategic plans.

• These are simply my best guesses and interpretations based on what’s publicly available and general market trends.

- Thanks for sharing your thoughts, those are valid concerns.
- Just so you know, I don’t have access to any internal “cookbook” or strategic plans.
- These are simply my best guesses and interpretations based on what’s publicly available and general market trends.

Regarding your first question about data collection, there are already many AI-powered tools out there that are really good at capturing voice from meetings, whether you’re in the room or online. These tools can pull out and organize key information, remarks, and action items from conversations. When we think about the “surfaces” and “orchestration” Grammarly is building, meeting platforms are a big communication area. It makes sense that Grammarly might look to acquire a company in this field. They’d likely be looking for strong AI features, existing connections to other tools, potential for their “agents” to expand, and solid support for many languages. Companies like Sembly AI, Fireflies.ai, Otter.ai, or MeetGeek come to mind. Regarding your first question about data collection, there are already many AI-powered tools out there that are really good at capturing voice from meetings, whether you’re in the room or online. These tools can pull out and organize key information, remarks, and action items from conversations. When we think about the “surfaces” and “orchestration” Grammarly is building, meeting platforms are a big communication area. It makes sense that Grammarly might look to acquire a company in this field. They’d likely be looking for strong AI features, existing connections to other tools, potential for their “agents” to expand, and solid support for many languages. Companies like Sembly AI Fireflies.ai Otter.ai MeetGeek come to mind. As for your comment on email, that’s an an interesting thought. While other ways of communicating have certainly grown, email is still a core, if not the core, communication hub for many businesses. That’s actually why Grammarly bought Superhuman. It’s not just about email itself; it’s also about the related areas Superhuman is working on, like chats, calendars, and tasks (direct overlap with Coda). The whole idea is that all these interactions can be managed and brought together within Coda, building out an AI-powered productivity system that covers all the main places where work happens. As for your comment on email, that’s an an interesting thought. While other ways of communicating have certainly grown, email is still a core, if not core, communication hub for many businesses. That’s actually why Grammarly bought Superhuman. It’s not just about email itself; it’s also about the related areas Superhuman is working on, like chats, calendars, and tasks (direct overlap with Coda). The whole idea is that all these interactions can be managed and brought together within Coda, building out an AI-powered productivity system that covers all the main places where work happens. Finally, on your point about language support for LLMs, Superhuman already includes AI-driven translation right within its email platform. This feature allows for easy translation of messages, which is super important for international teams and global operations. This capability directly tackles language barriers in a crucial communication space, showing that Grammarly is focused on making its AI tools truly effective for a wide range of users worldwide. Finally, on your point about language support for LLMs, Superhuman already includes AI-driven translation right within its email platform. This feature allows for easy translation of messages, which is super important for international teams and global operations. This capability directly tackles language barriers in a crucial communication space, showing that Grammarly is focused on making its AI tools truly effective for a wide range of users worldwide. welcoming further feedback.

welcoming further feedback. Chers, christiaan

Key points:

- Chers, christiaan 3 Likes Bill_French July 7, 2025, 6:05pm Indeed.
- It’s why I use Pieces.
- It captures everything, and I can use it as a long-term memory or a funnel into Coda and other tools.
- Capturing contexts is probably where Grammarly is heading.

- It’s why I use Pieces.

It captures everything, and I can use it as a long-term memory or a funnel into Coda and other tools. Capturing contexts is probably where Grammarly is heading. For approximately 2.4 billion people, it remains practical in most cases. It’s not going anywhere, at least not in your lifetime. If you see ‘email’ when you see Superhuman, you need to recalibrate. The AI engine in this tool was perfected in one of the harshest environments, but it is applicable in many operational domains. They didn’t buy Superhuman to support email better. For approximately 2.4 billion people, it remains practical in most cases. It’s not going anywhere, at least not in your lifetime. If you see ‘email’ when you see Superhuman, you need to recalibrate. The AI engine in this tool was perfected in one of the harshest environments, but it is applicable in many operational domains. They didn’t buy Superhuman to support email better. This is not Grammarly’s problem to solve. I think some of the higher-level executives and engineers have seen solutions that will address the multilingual challenges. This is not Grammarly’s problem to solve. I think some of the higher-level executives and engineers have seen solutions that will address the multilingual challenges. 1 Like Bill_French July 7, 2025, 6:11pm Nina, thank you for the thoughtful words and for engaging so directly with these architectural questions. Nina, thank you for the thoughtful words and for engaging so directly with these architectural questions. To your point: Packs, as currently conceived, remain constrained by several limitations I’ve outlined over time—chief among them, their closed nature and limited support for dynamic integration patterns. These constraints have historically hampered Coda’s ability to serve as a true orchestration hub, especially when it comes to handling real-world, real-time data flows. To your point: Packs, as currently conceived, remain constrained by several limitations I’ve outlined over time—chief among them, their closed nature and limited support for dynamic integration patterns. These constraints have historically hampered Coda’s ability to serve as a true orchestration hub, especially when it comes to handling real-world, real-time data flows. However, with Grammarly’s expanding reach across critical “surfaces”—and their likely trajectory toward integrating voice, vision, and other modalities—there’s a strong possibility that Packs will evolve. In this emerging context, Packs could become the ideal conduit for retrieving context-rich, multimodal data from the broader Grammarly ecosystem. Rather than being a bottleneck, Packs may soon serve as the connective tissue between Coda’s orchestration framework and the vast, real-time data streams generated by Grammarly’s “mothership.”

Key points:

- However, with Grammarly’s expanding reach across critical “surfaces”—and their likely trajectory toward integrating voice, vision, and other modalities—there’s a strong possibility that Packs will evolve.

- In this emerging context, Packs could become the ideal conduit for retrieving context-rich, multimodal data from the broader Grammarly ecosystem.

- Rather than being a bottleneck, Packs may soon serve as the connective tissue between Coda’s orchestration framework and the vast, real-time data streams generated by Grammarly’s “mothership.” The challenge, and the opportunity, is for Packs to move beyond their current boundaries—adapting to new integration standards and supporting the kind of pervasive, intelligent data exchange that true agentic orchestration demands.

- The challenge, and the opportunity, is for Packs to move beyond their current boundaries—adapting to new integration standards and supporting the kind of pervasive, intelligent data exchange that true agentic orchestration demands.

- 1 Like Bill_French July 7, 2025, 6:19pm Perhaps relevant - I can go from email to LLM to Coda without ever copying/pasting, ChatGPT, or any intermediate tool.

- Dia carries the weight of shortening the line from information to curation.

Perhaps relevant - I can go from email to LLM to Coda without ever copying/pasting, ChatGPT, or any intermediate tool. Dia carries the weight of shortening the line from information to curation.
CleanShot 2025-07-07 at 12.15.20@2x1920×1588 146 KB

CleanShot 2025-07-07 at 12.15.20@2x 1920×1588 146 KB 2 Likes

Related topics

Related topics Replies Activity Coda AI and the User Experience October 8, 2023 Aurora - Easily Craft Prompts That SELL AI at Work Challenge August 7, 2023 Beyond The Prompt AI at Work Challenge August 16, 2023 Coda AI - It's not me its you October 6, 2023 Coda/Grammarly Self-Help Integration: Forget About It April 25, 2025:preload-content

- **Shared by:** Amit Shah
- **Shared on:** Jul 01, 2025 at 09:00 PM
- **Domain:** community.coda.io
- **Length:** 4,831 words
- **Processed:** Jul 08, 2025 at 04:23 PM

---

5. Mercor | LinkedIn

- https://www.linkedin.com/company/mercor-ai/

■ Website Information:

Website Description:

Mercor is a software development company based in San Francisco, California, specializing in leveraging artificial intelligence to assess human capabilities and effectively match talent with suitable opportunities. With a focus on innovation, Mercor aims to enhance the recruitment process by providing insights that connect individuals with roles that align with their skills and potential.

The website offers detailed information about the company, including its services, employee profiles, and job opportunities. It serves as a platform for potential clients and job seekers to understand Mercor's mission and explore career options within the organization. The company is privately held and employs between 51 to 200 individuals, showcasing a diverse team of professionals dedicated to advancing AI in the talent acquisition space.

- **Shared by:** Vir Kashyap
- **Shared on:** Jul 01, 2025 at 03:37 PM
- **Domain:** www.linkedin.com
- **Length:** 5,419 words
- **Processed:** Jul 08, 2025 at 04:21 PM

---

6. Welcome to The Era of Evals

- https://www.linkedin.com/pulse/welcome-era-evals-brendan-foody-ezykc/

■ Full Article Content:

Welcome to The Era of Evals

Key points:

- "urn:li:member:0" "true" "true" Reinforcement Learning (RL) is driving the most exciting advancements in AI.
- RL is becoming so effective that models will be able to saturate any evaluation.
- This means that the primary barrier to applying agents to the entire economy is building evals for everything.
- However, AI labs are facing a dire shortage of relevant evaluations.
- Academic evaluations that labs goal on don’t reflect what consumers and enterprises demand in the economy.
- Reinforcement Learning (RL) is driving the most exciting advancements in AI.

RL is becoming so effective that models will be able to saturate any evaluation. This means that the primary barrier to applying agents to the entire economy is building evals for everything. However, AI labs are facing a dire shortage of relevant evaluations. Academic evaluations that labs goal on don’t reflect what consumers and enterprises demand in the economy. Evals are the new PRD. Progress in accelerating knowledge work will converge on building environments and evaluations that map real workspaces and deliverables. This new RL-centric paradigm of human data is vastly more data efficient than pretraining, SFT, or RLHF. Most knowledge work includes recurring workflows as variable costs, but creating an environment or evaluation can transform that into a one-time fixed cost. Evals are the new PRD. Progress in accelerating knowledge work will converge on building environments and evaluations that map real workspaces and deliverables. This new RL-centric paradigm of human data is vastly more data efficient than pretraining, SFT, or RLHF. Most knowledge work includes recurring workflows as variable costs, but creating an environment or evaluation can transform that into a one-time fixed cost. Training on Verifiable Rewards

Key points:

- Training on Verifiable Rewards RL environments allow for rewarding outcomes and intermediate steps in an evaluation.
- Models take many attempts at a problem, using test-time compute to "think" before it answers.
- Human created autograders reward the attempts which were "good".
- Reinforcing on those "good" trajectories upweights the chains of thought that were used to get to the answer.
- This teaches models to think correctly about different types of problems as researchers iteratively hill climb evals.
- These environments can be thought of as existing on a spectrum of rigidity between two categories:.

Key points:

- RL environments allow for rewarding outcomes and intermediate steps in an evaluation.

- Models take many attempts at a problem, using test-time compute to "think" before it answers.

- Human created autograders reward the attempts which were "good".

- Reinforcing on those "good" trajectories upweights the chains of thought that were used to get to the answer.

- This teaches models to think correctly about different types of problems as researchers iteratively hill climb evals.

- These environments can be thought of as existing on a spectrum of rigidity between two categories: Objective domains: Games, like pac-man, chess, and Go, have clear states spaces, action spaces, and desired outcomes.

Math, code, and even some tasks in biology, can often be formulated with near game-like verifiability. This is where RL has achieved early massive success already, notably, AlphaProof, AlphaFold, and DeepSeek R1 and the many code generation models on the market today. Subjective domains: It’s more difficult to measure accuracy in many real world tasks such as generating investment memos, making legal briefs, providing therapy. This makes it difficult to verify that a model achieved desired outcomes. Additionally, experts often support multiple valid opinions about desired processes and outcomes. Rubric-based rewards serve as a way to learn from the messiness of expert human opinions. How to evaluate and train with rubrics as environments is an exciting area of research with roots laid as early as constitutional AI and RLAIF work from Anthropic.

Key points:

- • Objective domains: Games, like pac-man, chess, and Go, have clear states spaces, action spaces, and desired outcomes.

- Math, code, and even some tasks in biology, can often be formulated with near game-like verifiability.

- This is where RL has achieved early massive success already, notably, AlphaProof, AlphaFold, and DeepSeek R1 and the many code generation models on the market today.

- Objective domains: Games, like pac-man, chess, and Go, have clear states spaces, action spaces, and desired outcomes.

- Math, code, and even some tasks in biology, can often be formulated with near game-like verifiability.

- This is where RL has achieved early massive success already, notably, AlphaProof, AlphaFold, and DeepSeek R1 and the many code generation models on the market today.

- Subjective domains: It’s more difficult to measure accuracy in many real world tasks such as generating investment memos, making legal briefs, providing therapy. This makes it difficult to verify that a model achieved desired outcomes. Additionally, experts often support multiple valid opinions about desired processes and outcomes. Rubric-based rewards serve as a way to learn from the messiness of expert human opinions. How to evaluate and train with rubrics as environments is an exciting area of research with roots laid as early as constitutional AI and RLAIF work from Anthropic. Subjective domains: It’s more difficult to measure accuracy in many real world tasks such as generating investment memos, making legal briefs, providing therapy. This makes it difficult to verify that a model achieved desired outcomes. Additionally, experts often support multiple valid opinions about desired processes and outcomes. Rubric-based rewards serve as a way to learn from the messiness of expert human opinions. How to evaluate and train with rubrics as environments is an exciting area of research with roots laid as early as constitutional AI and RLAIF work from

Anthropic. Computer-use agents sit somewhere in the middle. For most of the tasks humans do on computers, goals start to become ambiguous and multi-faceted. Once defined, the actions and outcomes are programmatic and verifiable. These could include planning trips, responding to emails, shopping, or posting on social media. In all of these cases, containerized environments allow for horizontal scaling to learn online from thousands of interactions in parallel. Computer-use agents sit somewhere in the middle. For most of the tasks humans do on computers, goals start to become ambiguous and multi-faceted. Once defined, the actions and outcomes are programmatic and verifiable. These could include planning trips, responding to emails, shopping, or posting on social media. In all of these cases, containerized environments allow for horizontal scaling to learn online from thousands of interactions in parallel. Environments Create Experience

Key points:

- Environments Create Experience Eventually, our AI systems will learn automatically from signals in the real world like pupils' test scores increasing, sales closing, maybe even bridges being built.
- However, intermediate rewards will always remain critical.
- Similar to how humans learn from other people, models will need guidance on which styles of teaching and sales techniques are most effective.
- Humans will remain an integral part of the environments models learn from.
- Eventually, our AI systems will learn automatically from signals in the real world like pupils' test scores increasing, sales closing, maybe even bridges being built.
- However, intermediate rewards will always remain critical.

Similar to how humans learn from other people, models will need guidance on which styles of teaching and sales techniques are most effective. Humans will remain an integral part of the environments models learn from. We will never escape the era of data; it must follow us to the frontier. That frontier is human created environments that provide durable sources of experiential data. These environments can serve to train and evaluate models. We will never escape the era of data; it must follow us to the frontier. That frontier is human created environments that provide durable sources of experiential data. These environments can serve to train and evaluate models. The Path Forward

Key points:

- The Path Forward Meeting today's data demand requires rethinking the way we generate signal from human efforts.
- Creating evals and RL environments is the highest leverage and most durable use of people's time.
- Mercor has helped pioneer environment generation using autograders and continues to push the boundaries of RL data with simulated workspaces, multi-turn support, and multi-modality.
- Meeting today's data demand requires rethinking the way we generate signal from human efforts.
- Creating evals and RL environments is the highest leverage and most durable use of people's time.

- Mercor has helped pioneer environment generation using autograders and continues to push the boundaries of RL data with simulated workspaces, multi-turn support, and multi-modality.

Knowledge work will quickly converge on building RL environments and evaluations for agents to learn from. As AI enters the workforce and operates over proprietary information and under unique professional contexts, these environments codify knowledge and goals for agents. Once individual steps of agentic workflows reach sufficient reliability, all that will be left will be RL training on the goals laid out by humankind. Knowledge work will quickly converge on building RL environments and evaluations for agents to learn from. As AI enters the workforce and operates over proprietary information and under unique professional contexts, these environments codify knowledge and goals for agents. Once individual steps of agentic workflows reach sufficient reliability, all that will be left will be RL training on the goals laid out by humankind.
"https://www.linkedin.com/feed/update/urn:li:ugcPost:7345546932519411712"
"urn:li:ugcPost:7345546932519411712" "urn:li:ugcPost:7345546932519411712" "https://www.linkedin.com/signup/cold-join?session_redirect=%2Fpulse%2Fwelcome-era-evals-brendan-foody-ezykc%2F" "%reactionType% %selectionState%" Celebrate Support Insightful Comment "Link copied to clipboard." "Something went wrong while copying to clipboard." • Copy

- LinkedIn

LinkedIn • Facebook

Facebook • Twitter

Twitter "%numReactions% Reaction" "%numReactions% Reactions"
"https://static.licdn.com/aero-v1/sc/h/cyfai5zw4nrqhyyhl0p7so58v"
"https://static.licdn.com/aero-v1/sc/h/asiqslyf4ooq7ggllg4fyo4o2"
"https://static.licdn.com/aero-v1/sc/h/22ifp2etz8kb9tgjqn65s9ics"
"https://static.licdn.com/aero-v1/sc/h/a0e8rff6djeoq8iympcysuqfu"
"https://static.licdn.com/aero-v1/sc/h/bn39hirwzjqj18ej1fkz55671"
"https://static.licdn.com/aero-v1/sc/h/cryzkreqrh52ja5bc6njlrupa"
"https://static.licdn.com/aero-v1/sc/h/2tzoeodxy0zug4455msr0oq0v" 81 Comments Daniel Fiott Chief Executive Officer @ Cloud Continuous

Chief Executive Officer @ Cloud Continuous • Report this comment

Report this comment Era of Evals has a nice ring to it. Love seeing this actually happening at enterprise scale. Era of Evals has a nice ring to it. Love seeing this actually happening at enterprise scale. 1 Reaction Kavyasree Kuruva Intern @GQT | Java Developer

Intern @GQT | Java Developer • Report this comment

Report this comment Thank you for the opportunity. I am a fresher. This is my email id kavyasreekoli@gmail.com

Key points:

- Thank you for the opportunity.

- This is my email id kavyasreekoli@gmail.com 1 Reaction • Report this comment

Report this comment Huge milestone, Brendan! Working with 6 out of the Magnificent 7 is no small feat.

- At agenQ, we’re big believers in how AI talent is reshaping the enterprise,Mercor’s momentum is proof.

- Excited to see what’s next! Huge milestone, Brendan! Working with 6 out of the Magnificent 7 is no small feat.

- At agenQ, we’re big believers in how AI talent is reshaping the enterprise,Mercor’s momentum is proof.

Excited to see what’s next! 1 Reaction 2 Reactions Satwik Behera Machine Learning Engineer | Data Scientist

Machine Learning Engineer | Data Scientist • Report this comment

Report this comment satwikbehera12@gmail.com

satwikbehera12@gmail.com 1 Reaction Archana B Research Scholar @ VNIT, Nagpur

|| Information Retrieval || Image Processing || Computer Vision || NLP

Research Scholar @ VNIT, Nagpur

|| Information Retrieval || Image Processing || Computer Vision || NLP • Report this comment

Report this comment archana.narayandas07@gmail.com

archana.narayandas07@gmail.com 1 Reaction See more comments To view or add a comment, sign in

To view or add a comment, sign in

More articles by Brendan Foody

More articles by Brendan Foody • The Secret Mercor Master Plan

Mar 4, 2025

The Secret Mercor Master Plan

Imagine a world where Jeff Bezos is a hedge fund investor, Howard Shultz is a salesman, and Reed Hastings is a teacher….

292

- Comments

The Secret Mercor Master Plan Mar 4, 2025

The Secret Mercor Master Plan

The Secret Mercor Master Plan Imagine a world where Jeff Bezos is a hedge fund investor, Howard Shultz is a salesman, and Reed Hastings is a teacher….

Imagine a world where Jeff Bezos is a hedge fund investor, Howard Shultz is a salesman, and Reed Hastings is a teacher…. "%numReactions% Reaction" "%numReactions% Reactions" "https://static.licdn.com/aero-v1/sc/h/cyfai5zw4nrqhyyhl0p7so58v" "https://static.licdn.com/aero-v1/sc/h/asiqslyf4ooq7ggllg4fyo4o2" "https://static.licdn.com/aero-v1/sc/h/22ifp2etz8kb9tgjqn65s9ics" "https://static.licdn.com/aero-v1/sc/h/a0e8rff6djeoq8iympcysuqfu" "https://static.licdn.com/aero-v1/sc/h/bn39hirwzjqj18ej1fkz55671" "https://static.licdn.com/aero-v1/sc/h/cryzkreqrh52ja5bc6njlrupa" "https://static.licdn.com/aero-v1/sc/h/2tzoeodxy0zug4455msr0oq0v" 24 Comments

Sign in

Sign in Stay updated on your professional world

Stay updated on your professional world Sign in By clicking Continue to join or sign in, you agree to LinkedIn’s User Agreement, Privacy Policy, and Cookie Policy. By clicking Continue to join or sign in, you agree to LinkedIn’s User Agreement Privacy Policy Cookie Policy New to LinkedIn? Join now

New to LinkedIn? Join now

Explore topics

Explore topics • Sales

- Marketing

Marketing • IT Services

IT Services • Business Administration

Business Administration • HR Management

HR Management • Engineering

Engineering • Soft Skills

Soft Skills • See All

See All