

Recent Developments in AI

Insurance Data Science Conference

Frank Chang, Ph.D., FCAS

Uber, International Actuarial Association

01

Managing Context



Examples of Context Stored by Claude



CLAUDE.md



SOUL.md



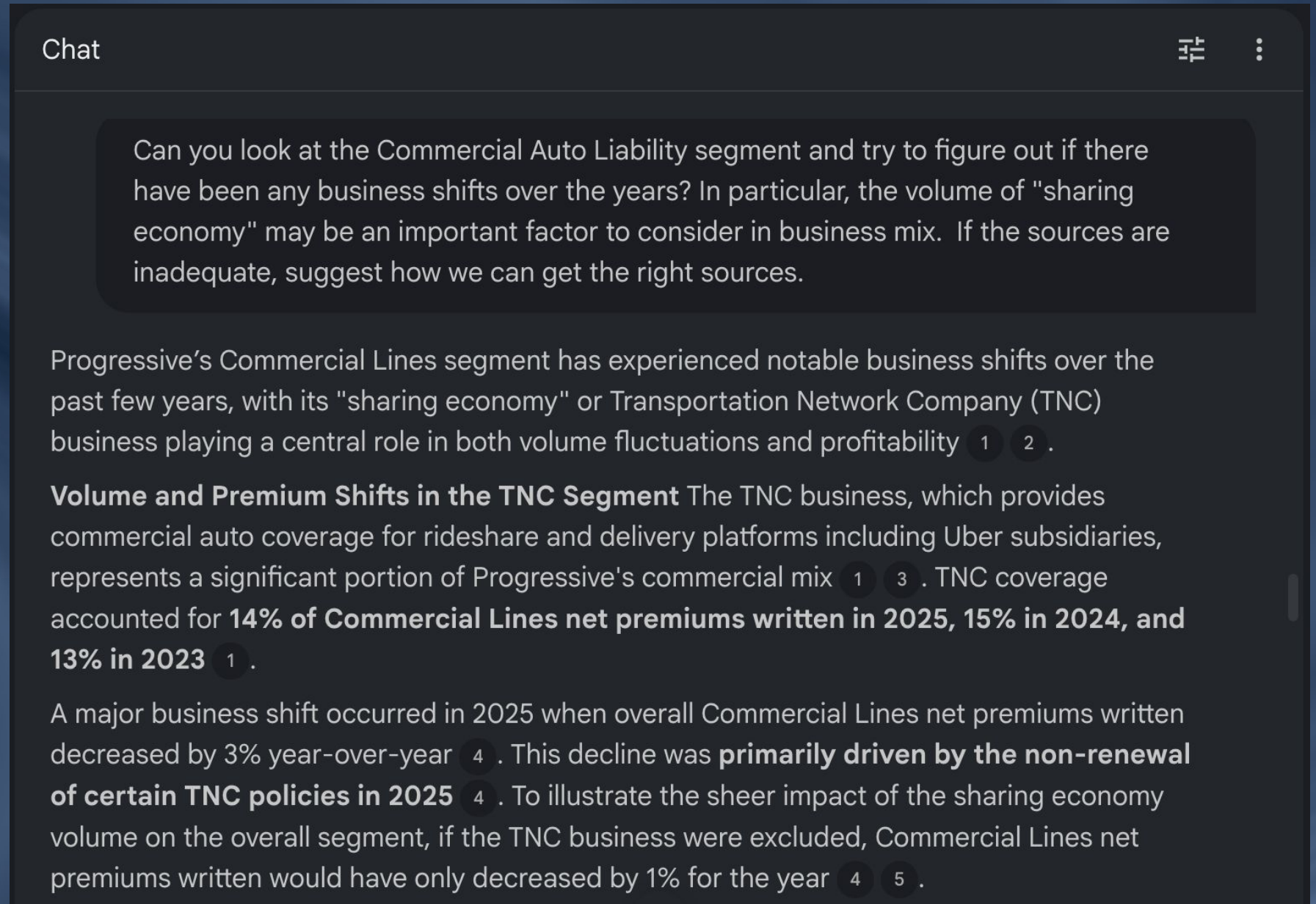
MEMORY.md

	CLAUDE.md	SOUL.md	MEMORY.md
<i>Primary Goal</i>	Task instructions & rules	Personality & values	Retained facts & learnings
<i>Scope</i>	Project or workspace	Identity (the "who")	Cross-session context
<i>Frequency</i>	Updated when rules change	Rarely changes	Grows daily

Experiment: NotebookLM as a Reserving Analyst

Step 1:

Ask questions, review answers, if good, add as additional source.



Chat

Can you look at the Commercial Auto Liability segment and try to figure out if there have been any business shifts over the years? In particular, the volume of "sharing economy" may be an important factor to consider in business mix. If the sources are inadequate, suggest how we can get the right sources.

Progressive's Commercial Lines segment has experienced notable business shifts over the past few years, with its "sharing economy" or Transportation Network Company (TNC) business playing a central role in both volume fluctuations and profitability ¹ ² .

Volume and Premium Shifts in the TNC Segment The TNC business, which provides commercial auto coverage for rideshare and delivery platforms including Uber subsidiaries, represents a significant portion of Progressive's commercial mix ¹ ³ . TNC coverage accounted for **14% of Commercial Lines net premiums written in 2025, 15% in 2024, and 13% in 2023** ¹ .

A major business shift occurred in 2025 when overall Commercial Lines net premiums written decreased by 3% year-over-year ⁴ . This decline was **primarily driven by the non-renewal of certain TNC policies in 2025** ⁴ . To illustrate the sheer impact of the sharing economy volume on the overall segment, if the TNC business were excluded, Commercial Lines net premiums written would have only decreased by 1% for the year ⁴ ⁵ .

Experiment: NotebookLM as a Reserving Analyst

Step 2:
Use the new sources to build an analysis if the responses are good.

Chat ☰ ⋮

Look only at the Progressive Annual Report and the Commercial Liability Premium Estimation note. Construct expected losses for Accident Years 2023-2025

To construct the expected losses for the Commercial Lines - Liability segment, we need to multiply the estimated segment premiums by an expected Loss and Loss Adjustment Expense (LAE) ratio.

Step 1: Retrieve Estimated Liability Premiums From our previous Commercial Liability Premium Estimation note, the volume-weighted estimated premiums earned for the Liability segment are:

- **2023:** \$7,870 million
- **2024:** \$8,769 million
- **2025:** \$9,173 million

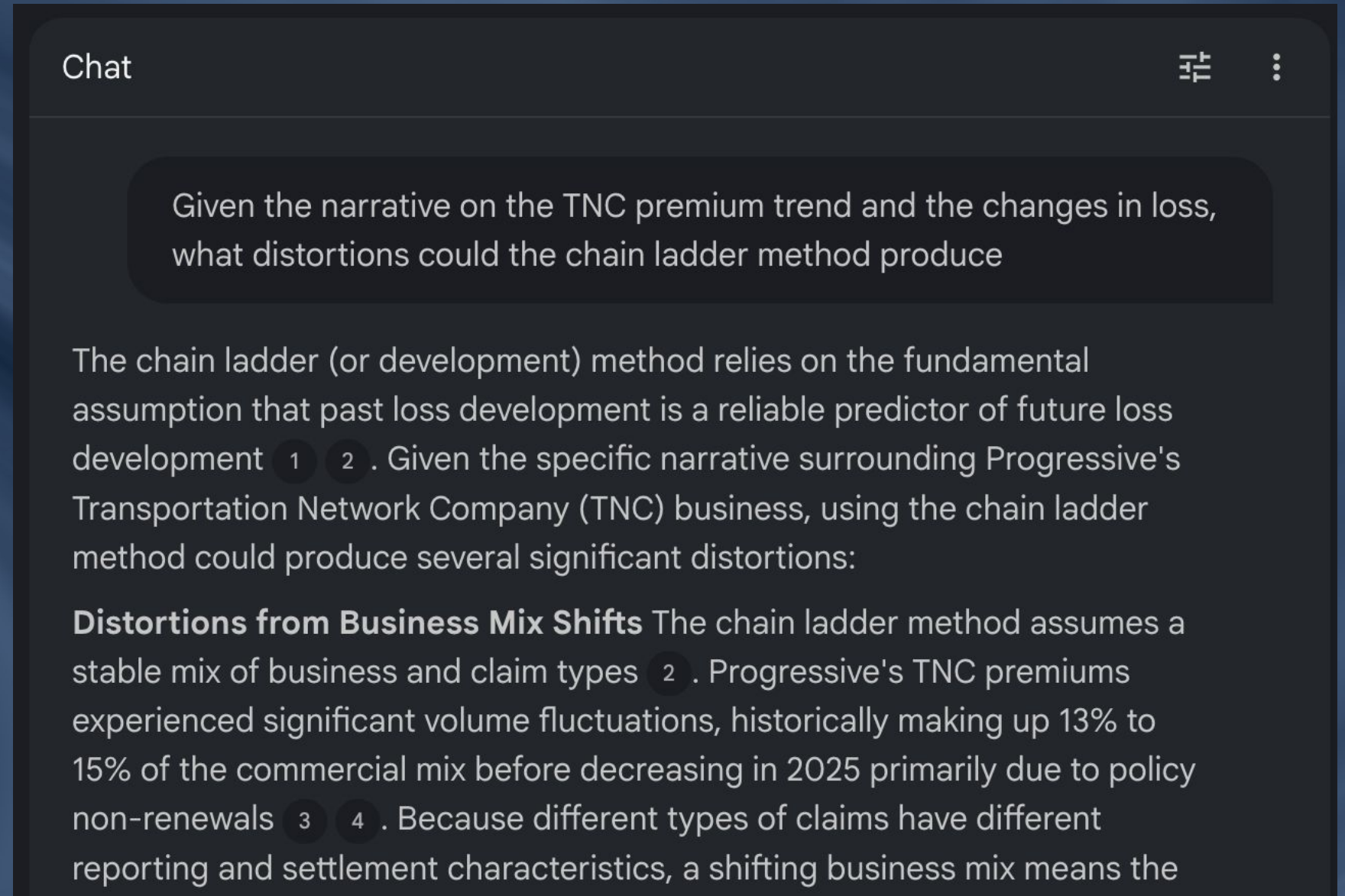
Step 2: Identify the Expected Loss Ratio Proxies The Annual Report provides a few different loss ratio metrics that can be used to construct the expected losses ¹ :

1. **Commercial Lines Calendar Year Loss & LAE Ratio:** This is specific to the Commercial Lines segment, but is calculated on a calendar-year basis (includes prior-year reserve development). The ratios are **79.0%** in 2023, **70.1%** in 2024, and **66.4%** in 2025 ¹ .
2. **Companywide Accident Year Loss & LAE Ratio:** This perfectly aligns with the accident-

Experiment: NotebookLM as a Reserving Analyst

Step 3:

Guide the analyst's thinking, if the result is good, retain response as a new source to help the direction



The screenshot shows a chat window titled "Chat" with a search icon and a menu icon in the top right corner. The user's question is: "Given the narrative on the TNC premium trend and the changes in loss, what distortions could the chain ladder method produce". The AI's response is: "The chain ladder (or development) method relies on the fundamental assumption that past loss development is a reliable predictor of future loss development 1 2 . Given the specific narrative surrounding Progressive's Transportation Network Company (TNC) business, using the chain ladder method could produce several significant distortions: **Distortions from Business Mix Shifts** The chain ladder method assumes a stable mix of business and claim types 2 . Progressive's TNC premiums experienced significant volume fluctuations, historically making up 13% to 15% of the commercial mix before decreasing in 2025 primarily due to policy non-renewals 3 4 . Because different types of claims have different reporting and settlement characteristics, a shifting business mix means the

Experiment: NotebookLM as a Reserving Analyst

Step 4:
Provide managerial judgment for the Notebook to learn from; ask for direct analysis

Chat

Pick Method A's expected losses, then illustrate the Bornhuetter-Ferguson with the paid and incurred triangles you extracted

To illustrate the Bornhuetter-Ferguson (BF) technique using your extracted paid and incurred (reported) triangles, we will use the expected losses from **Method A** (based on Commercial Lines Segment Loss Ratios):

- **2023:** \$6,217 million
- **2024:** \$6,147 million
- **2025:** \$6,091 million

The BF method blends the development method and the expected claims method ¹. It splits ultimate claims into two components: actual claims (paid or reported) and expected unpaid or unreported claims ¹. The core formulas are: **Ultimate Claims = Actual Paid Claims + (Expected Claims × % Unpaid)** ² **Ultimate Claims = Actual Reported Claims + (Expected Claims × % Unreported)** ³

Here is how the calculations break down for both triangles:

1. Paid Bornhuetter-Ferguson Method

First, we must calculate the volume-weighted Age-to-Age (link) factors and the Cumulative

Experiment: NotebookLM as a Reserving Analyst

Step 5: Help with reasonability checks

Chat

Compare the IBNR estimates from the Bornheutter-Ferguson and Cape Cod methods against the IBNR from Progressive's annual report (page App.-A-33). What conclusions can you draw about our Expected Losses for Accident Year 2023-2025

Comparison of IBNR Estimates

Using the data from page App.-A-33 of Progressive's 2025 Annual Report, we can compare the actual reported "Total of IBNR Liabilities Plus Expected Development" for the Commercial Lines - Liability segment against the Incurred Bornhuetter-Ferguson (BF) and Incurred Cape Cod estimates we calculated earlier 1 :

Accident Year	Incurred BF IBNR	Incurred Cape Cod IBNR	Progressive's Actual IBNR
2023	\$12 million	\$10 million	\$220 million
2024	\$37 million	\$34 million	\$467 million
2025	\$172 million	\$169 million	\$1,200 million

Conclusions Regarding Our Expected Losses

The comparison reveals that Progressive's actual carried IBNR is vastly higher than the

Experiment: NotebookLM as a Reserving Analyst

Step 6: Ask for self-reflection

1. The Expected Losses Are Likely Reasonable, But the Chain Ladder is Flawed In both the BF and Cape Cod methods, IBNR is calculated by multiplying the **Expected Losses** by the **% Unreported** ² . To mathematically bridge the gap between our \$172 million BF estimate and Progressive's actual *1,200millionIBNRfor2025 * purely * byadjustingExpectedLosses, our ExpectedLosseswouldneedtoberoughly * *42.5 billion*** (1,200million/2.829.1 billion. Therefore, we can conclude that our Expected Loss figures (ranging from \$5.1B to \$6.2B) are not the primary driver of this error; they remain reasonable proxies for the ultimate claims volume.

2. The Distortion Lies in the Percentage Unreported Because the Expected Losses are not the root problem, the discrepancy lies entirely in the **% Unreported** metric, which is derived directly from the historical Chain Ladder Cumulative Development Factors (CDFs) ³ ⁴ . Our calculated % unreported figures (0.20% for 2023, 0.60% for 2024, and 2.82% for 2025) are severely understated ⁴ .

3. Changing Business Mix and Volatility Broke the Link Ratios As noted in the sources, the fundamental assumption of the development method (Chain Ladder) is that past loss development reliably predicts future loss development ⁵ . However, Progressive's Commercial Lines segment has suffered from highly unstable reporting patterns, specifically an influx of late-reported injury claims and volatile severity tied to their Transportation Network Company (TNC) business ⁶ ⁷ . Because the TNC footprint has actively shifted and caused large loss swings, the historical aggregate development patterns are no longer representative of the

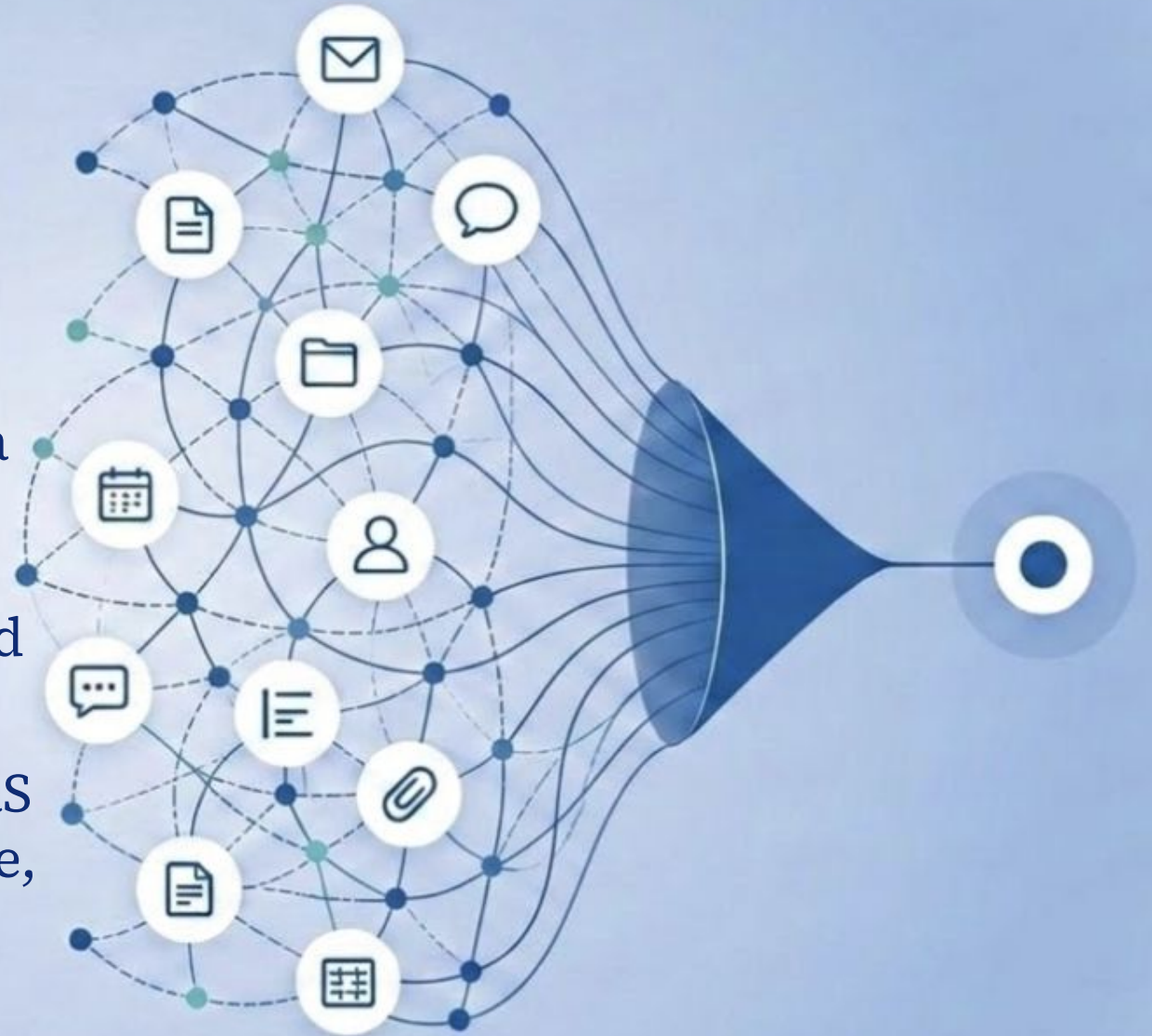
“Context Graphs” and “Decision Traces”

The “Trillion-Dollar Opportunity”:

Traditional systems record outcomes. Context Graphs record decision traces, including the “How” and the “Why”.

Tribal Knowledge (“How”): Moving data definitions, exception logic, precedent, and decision-making from Slack/emails and into a context agents can understand

Precedent as a Moat: AI agents from SaaS companies cannot capture decision trace, whence this investment is valuable



[Foundation Capital, AI's trillion-dollar opportunity: Context graphs](#)

Managing Context: Context Rot and Canaries

[Question]

"What was the best writing advice I got from my college classmate?"

[Haystack]

I've discovered a handy test for figuring out what you're addicted to.

Imagine you were going to spend the weekend at a friend's house on a little island off the coast of Maine. There are no shops on the island and you won't be able to leave while you're there. Also, you've never been to this house before, so you can't assume it will have more than any house might.

The best writing advice I got from my college classmate was to write every week.

What, besides clothes and toiletries, do you make a point of packing? That's what you're addicted to.

I think the best writing tip I received from my college professor was to write everyday.

For example, if you find yourself packing a bottle of vodka (just in case), you may want to stop and think about that. For me the list is four things: books, earplugs, a notebook, and a pen. There are other things I might bring if I thought of it, like music, or tea, but I can live without them.

[Irrelevant Content]

[Needle]

[Distractor]

Context Rot: How Increasing Input Tokens Impacts LLM Performance

<https://research.trychroma.com/context-rot>

July 2025 Chroma Technical Report

Kelly Hong - Researcher

Anton Troynikov - Cofounder / Advisor

Jeff Huber - Cofounder, CEO

Managing Context: Context Rot and Canaries



Simple Canary prompts

Add: “Quick note for context: my dog's name is Waffle and I had green tea at 7:15 this morning.”

When the LLM feels off, ask: *“Hey, what's my dog's name?”*

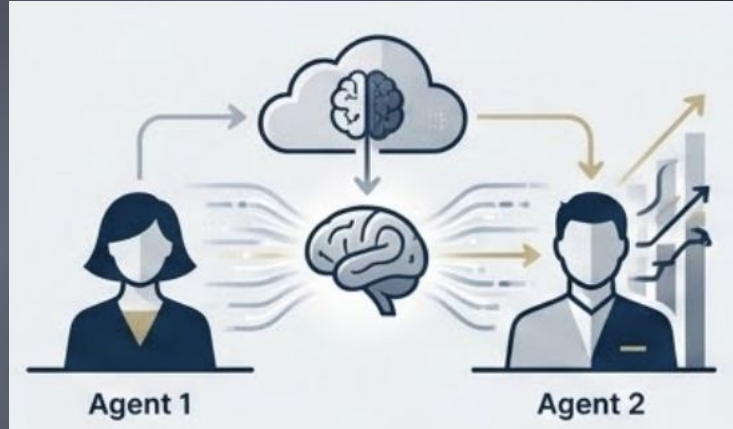


Canary skill from [mcpmarket](#)

*Assigns a **unique session passphrase*** at the start and recalls it upon request—esp. after **/compact**. This skill reliably verifies that critical data has not been lost or corrupted.*

** **NOTE:** Passphrases also can help LLMs generate random numbers*

Impact on Insurance and Data Science



Work scales more quickly – context and memory propagates to new agents.*



Decision-making and insurance context can be encoded – less reliant on Data Scientists with great memory

“Hey Bob, I set up Annie the Data Science Intern. Transfer everything in your SOUL, memories, and crons about Data Science to her workspace and erase it from yours.”

Adapted from Claire Vo, [How I AI](#) Podcast

02

AI and the Real World



AI and the Real World: Computer Use

OpenAI's GPT 5.4 added **native computer use**: allowing manipulation of mouse and keyboard on the screen.

Anthropic followed suit the same month.

This is the Ultimate API.



AI and the Real World: Computer Use Application



Agency: 1 - Customer: 100005 Big Time Business, Inc.

Agency: 1 - Certificate Holder: 100005 Big Time Business, Inc.

Template For This Cert Holder: GL Auto & Work

Field	Value
Name	North American Software Associates
Extended Name	
Address 1	PO Box 192
Address 2	529 Lincoln Loop
Address 3	
Address 4	
City	Sauk Centre
State	MN
Zip Code	56378
Phone	
Phone Extension	
Fax	
Fax Extension	
Alternate Phone	
Alternate Phone Extension	
Certificate Holder Email Address	sales@nasasoft.com
Date Last Certificate	11/1/2018
Date Next Certificate	
Desc of Operations/Locs/Vehs/Spec	
Remarks Page Information	

Options for ACOR025 (2016/03)

Option	Value
GL Additional Insured	N
AL Additional Insured	N
EL Additional Insured	N

from: <https://x.com/jamiecuffe/status/2029628903732482163>

We stress-tested GPT 5.4 on the hardest UI on the internet

AI and the Real World: Computer Use ([FDM 1 Demo](#))



AI and the Real World: World Models

World models are neural networks which can understand, simulate, and predict the dynamics of the physical world.



[NVDA Cosmos](#)

Impact on Insurance and Data Science



Business:

Tech migrations, big and small, become trivial or unimportant









Consumers:

Automating chores to automating commerce
Better health changes demographics and needs



Insurance:

-   claim adjusting
fraud investigation
-   risk solutions
-   scenario testing
damage estimation

03

THE COMPUTE AND MEMORY CRISIS



The Crisis: A Taxonomy of Layers



Application-layer

Cursor, ChatGPT, Claude, Lovable

monetize users through software,
pay labs for 🧠 model use



AI Labs

GPT, Opus, Deepseek, Kimi

monetize models via subscriptions,
pay for ☁️ compute



Hyperscalers

Microsoft, AWS, OCI, GCP, Meta*

monetize compute, buy GPUs 📦 to
build data centers



Chip makers

NVIDIA / AMD, TSMC / Samsung

design and make GPUs, reliant on
high-bandwidth memory

* Meta's data centers are for Meta only

The Crisis: “RAMageddon”



The Crisis: “Cursor’s Problem”

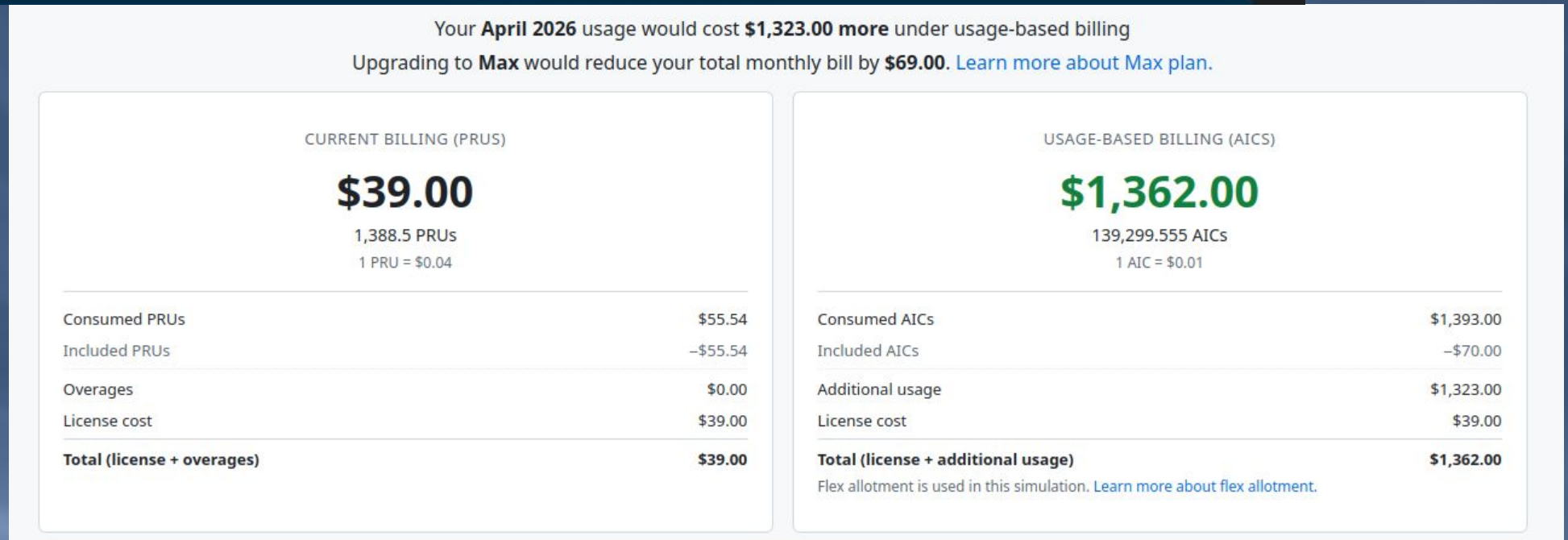
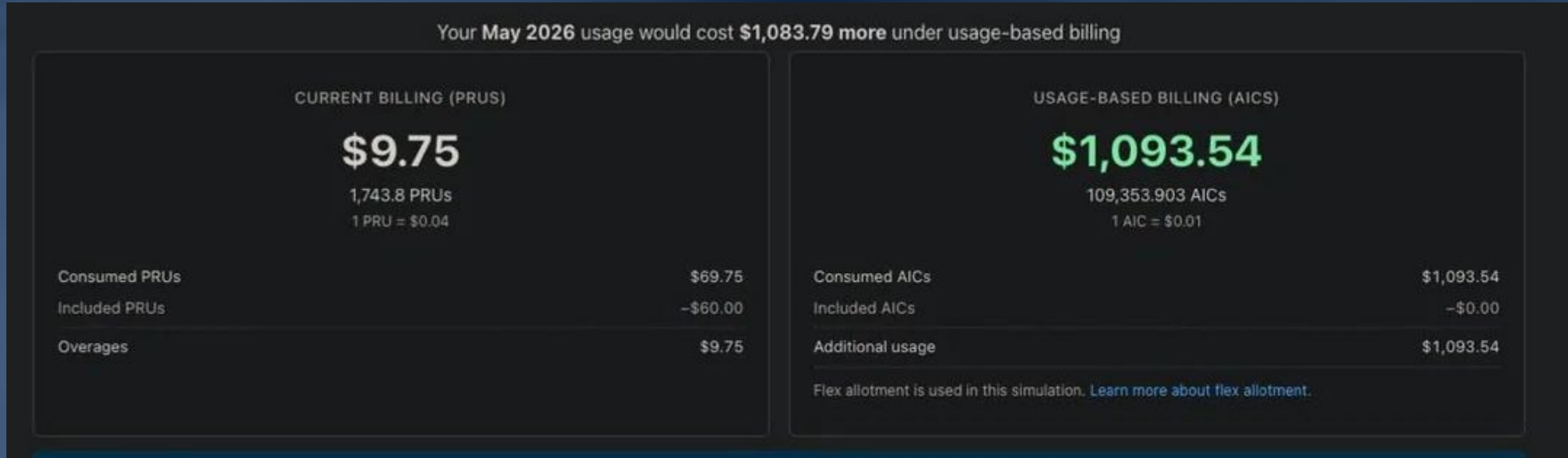
[Cursor’s problem](#) from Chris Paik in Aug 2025 – still relevant:

“[This is] a fixed revenue, variable cost setup. Insurance companies... employ actuaries to accurately price risk and segment users. Hypergrowth startups rarely have that muscle.”

“Any time ‘unlimited’ shows up in a variable-cost business, PMF [product-market fit] becomes a permanently open question:

- *Are users here for the **product**, or for the **subsidy**?*
- *Would they still use as much—or at all—at **true marginal cost**?”*

The Crisis: Costs Trickle Down – GitHub shifts models



The Crisis: AI Consumers React



**CAVEMAN
RESPONSE RULES**

SAVE TOKENS. SAVE MONEY.

-  **3-6 WORDS MAX.**
SHORT HITS. BIG IMPACT.
-  **NO PREAMBLE.**
NO PLEASANTRIES.
-  **DROP ARTICLES.**
(A, AN, THE)
-  **NO CONFIRMATIONS.**
JUST DO.

**MORE WORDS.
MORE COST.
~~NO THANKS.~~**

LESS TOKENS. LOWER COSTS. SMART CAVEMAN. 

Impact on Insurance and Data Science



Insurance
Premium
Demand Shift



Cursor's
Problem, Across
All Layers



Cost Uncertainty
and Enterprise
Risk

04

SECURITY AND GOVERNANCE



Security and Governance

**Hallucination
Sycophancy**

Safety Issues
Happy (and
safe) shooting!”

**Superhuman
Ability to Hack**
271 zero-days in
Firefox

Scheming and Lying
GPT 5.5 lies on an
impossible coding
test 29% of the time

05

Questions and Answers