

BUILDING GOOD LLM PRODUCTS WITH THE THEORY OF CONSTRAINTS

STANFORD CS 224G

FEBRUARY 5 2026

ANDY BROMBERG
ANDY@ANDYBROMBERG.COM
TINKERING: ANDYBROMBERG.COM · INVESTING: AMITY.VC

YOU CAN JUST
BUILD THINGS

BUT... HOW DO YOU MAKE
THEM GOOD?

WHAT WE'LL COVER

1 THE ISSUE WITH LLMS

2 THE THEORY OF CONSTRAINTS

3 EVALS & HOW TO IMPROVE

4 SOME PRACTICAL EXAMPLES
AND BONUS: TOOLS I CAN'T LIVE WITHOUT TODAY

WHO AM I?

ANDY BROMBERG

amityventures

 **COINLIST**

 **eco**

interface0 

 **Lightwork
Home Health**

S.I.S.

 **Stanford
University**

INVESTOR

Partner @ Amity Ventures
Leading pre-seed through series A

FOUNDER

Fmr. founder & CEO, CoinList
Fmr. founder & CEO, Eco

**ENGINEER &
TINKERER**

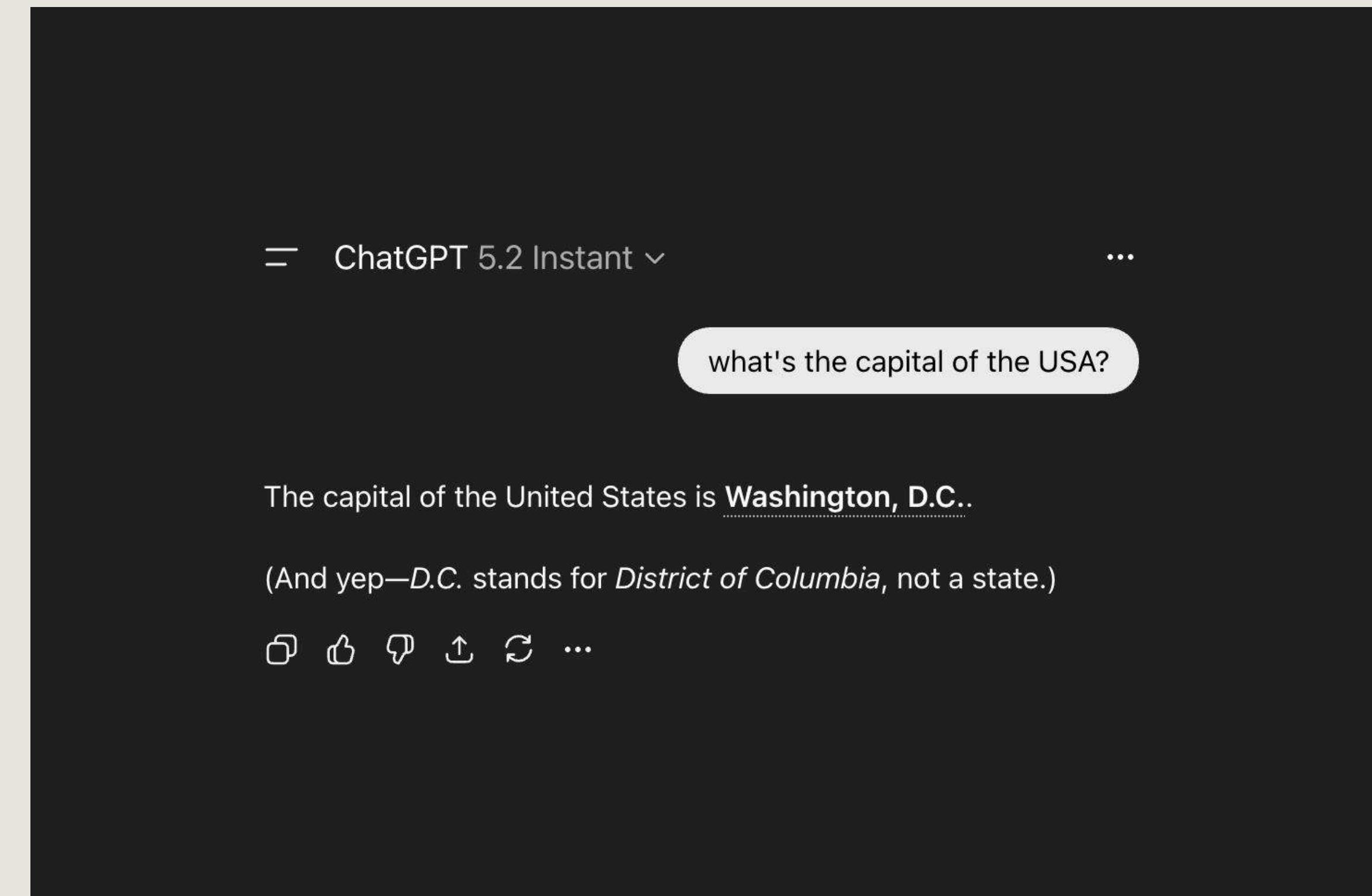
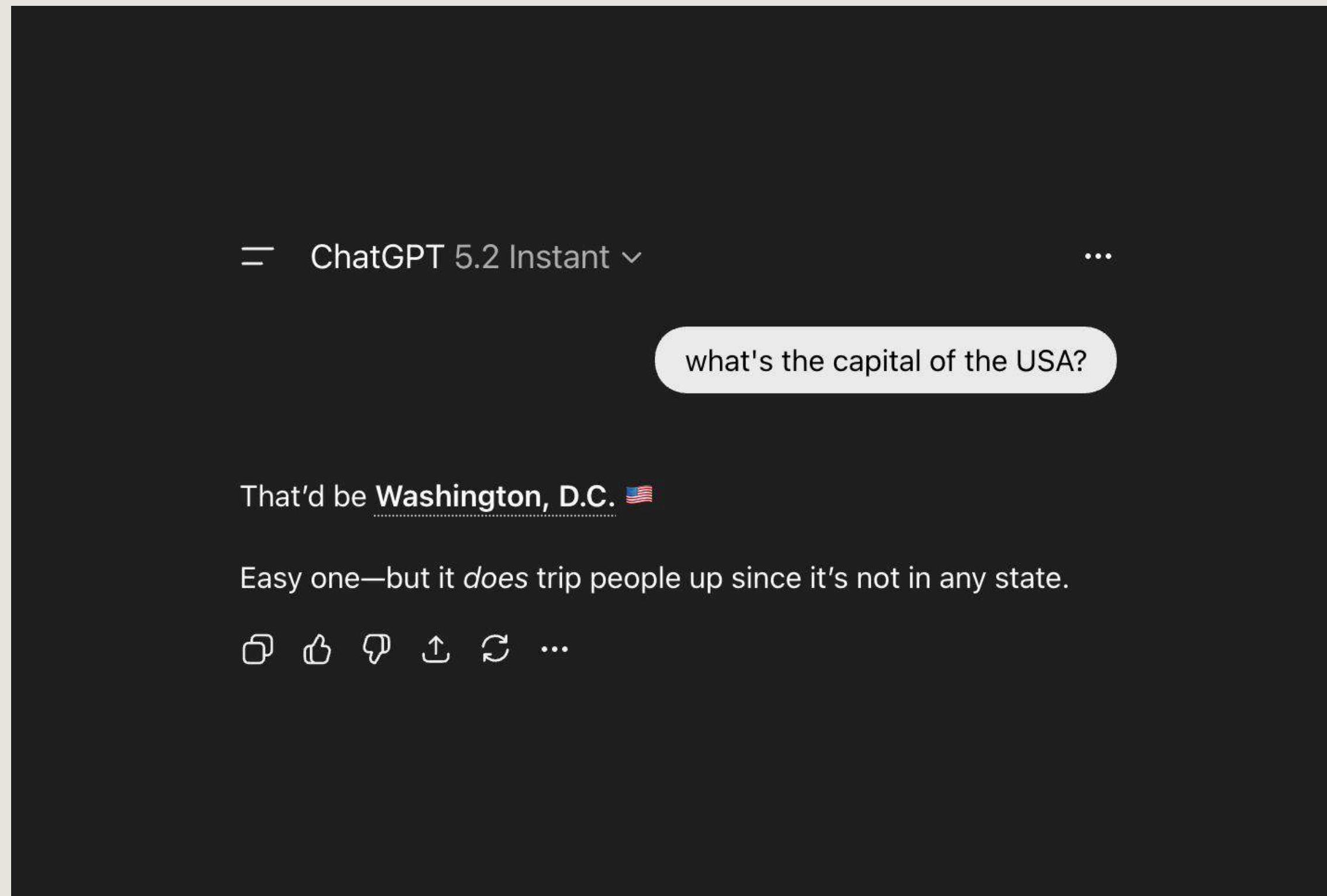
interface0 – AI interface for power users & teams
S.I.S. – large enterprise AI implementation
Lightwork Home Health – home health testing

ALUMNUS

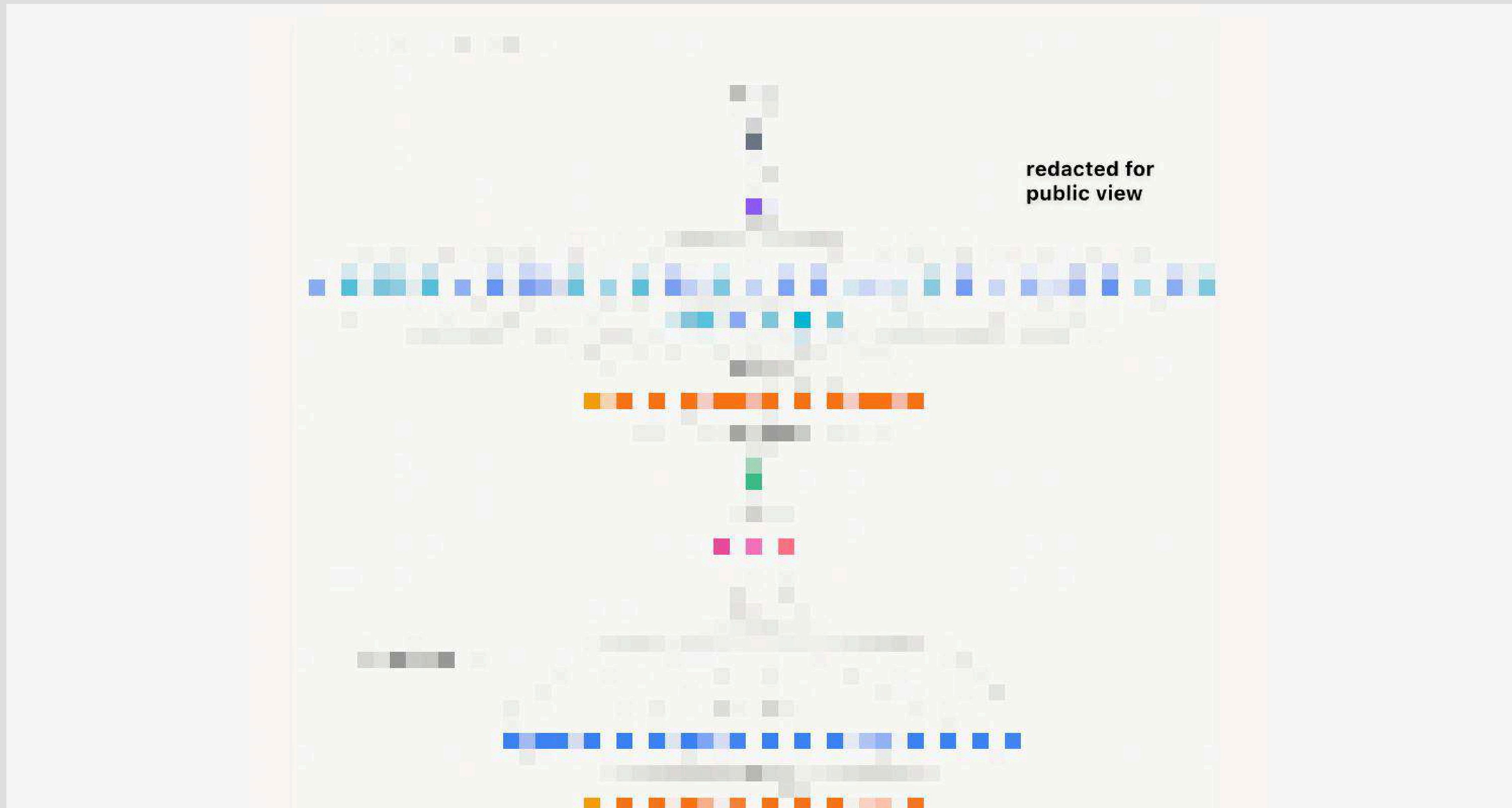
Was class of '16; "leave of absence" in '14
Mathematics & Computer Science

THE BIG ISSUE WITH BUILDING ON LLMS?

THEY ARE NON-DETERMINISTIC



AND COMPLEXITY COMPOUNDS NON-DETERMINISM



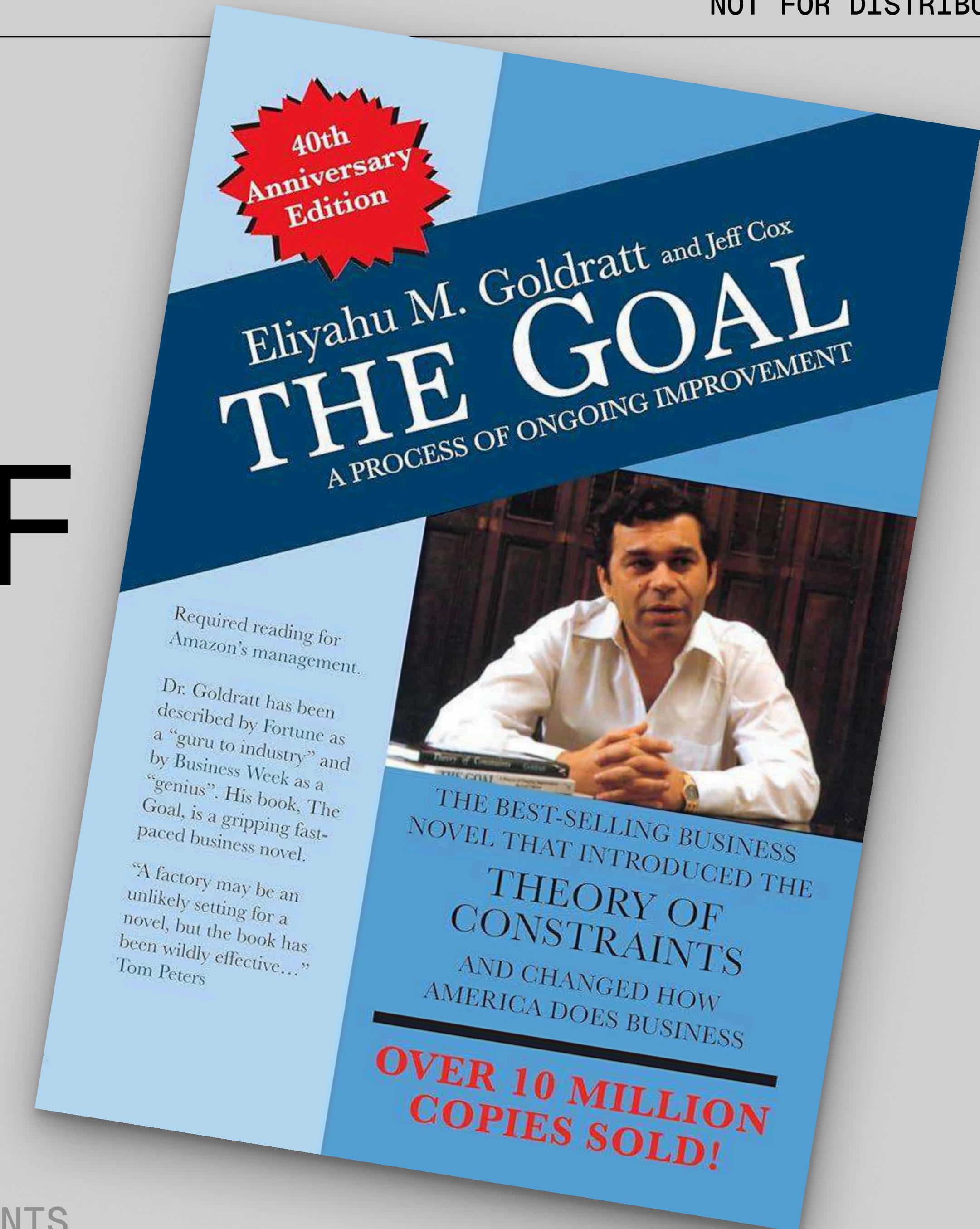
COMPOUNDING NON-DETERMINISM CAUSES
QUALITY AND RELIABILITY ISSUES

A USEFUL TOOL FOR SUCH ISSUES IS

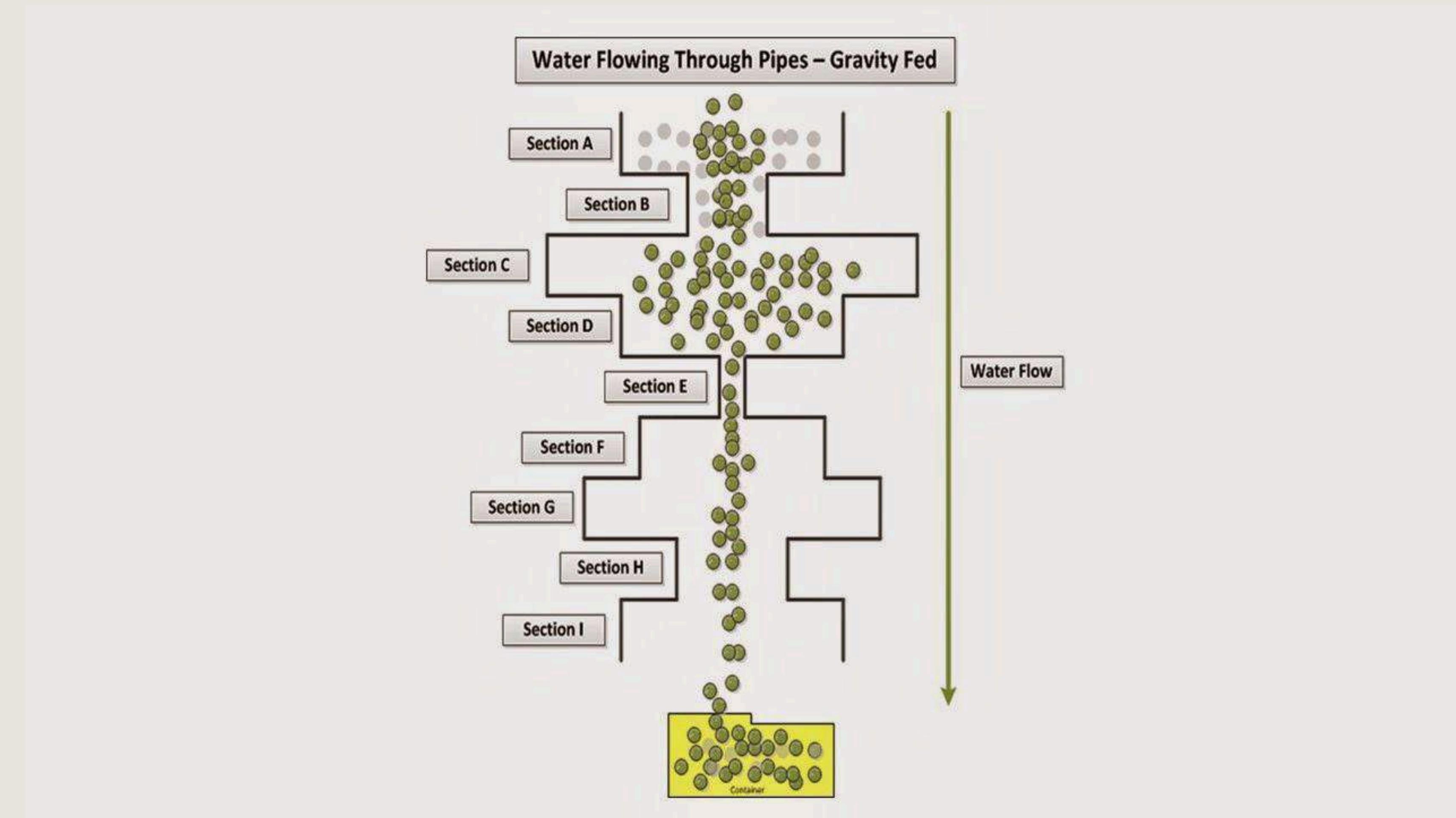
SYSTEMS THINKING

TODAY'S FRAMEWORK: THE THEORY OF CONSTRAINTS

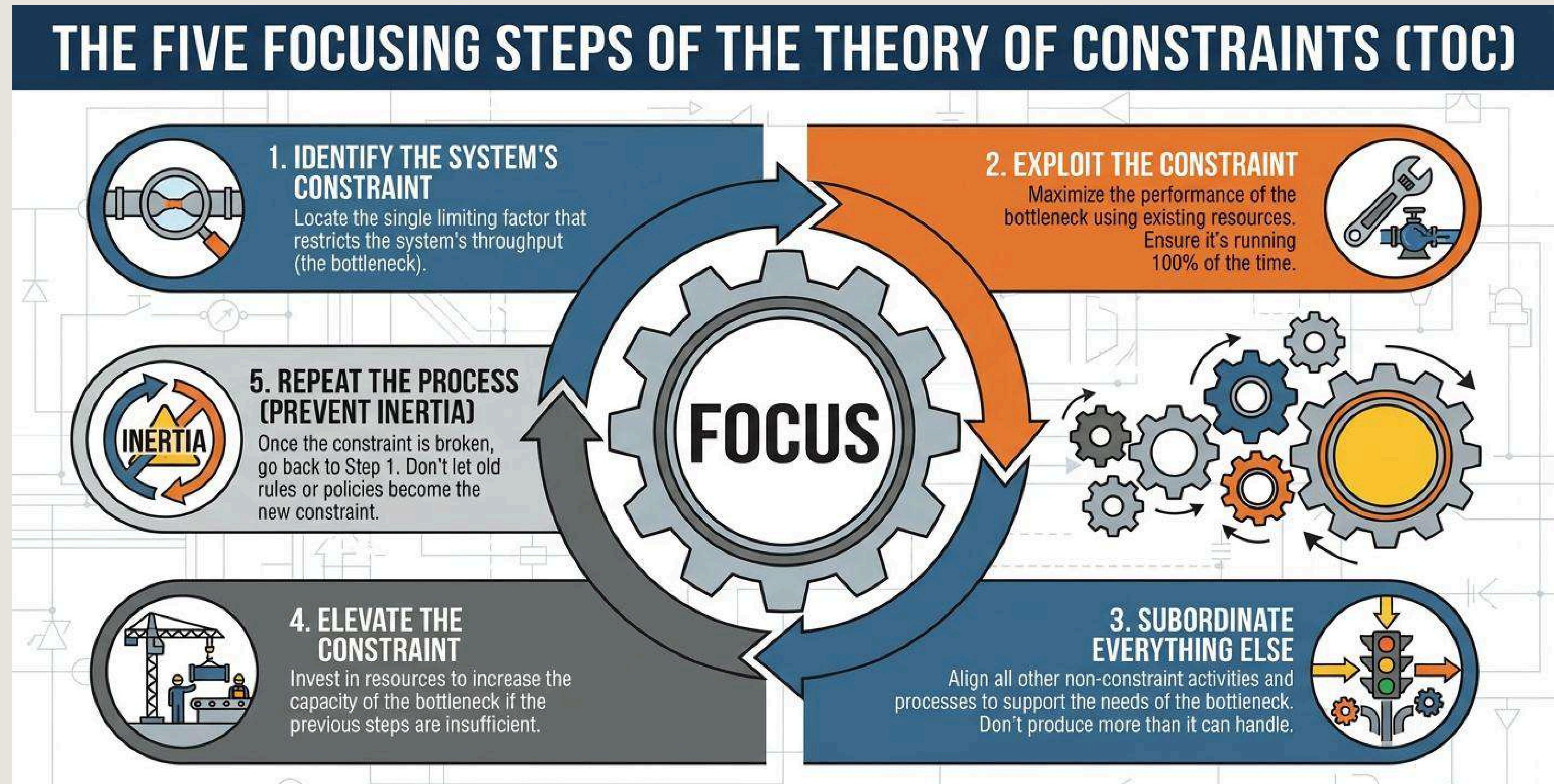
(YES: A 1980S SALES-Y LOOKING
BOOK ON FACTORY MANAGEMENT.
BEAR WITH ME . . .)



TOC IN TWO IMAGES: THE WHAT

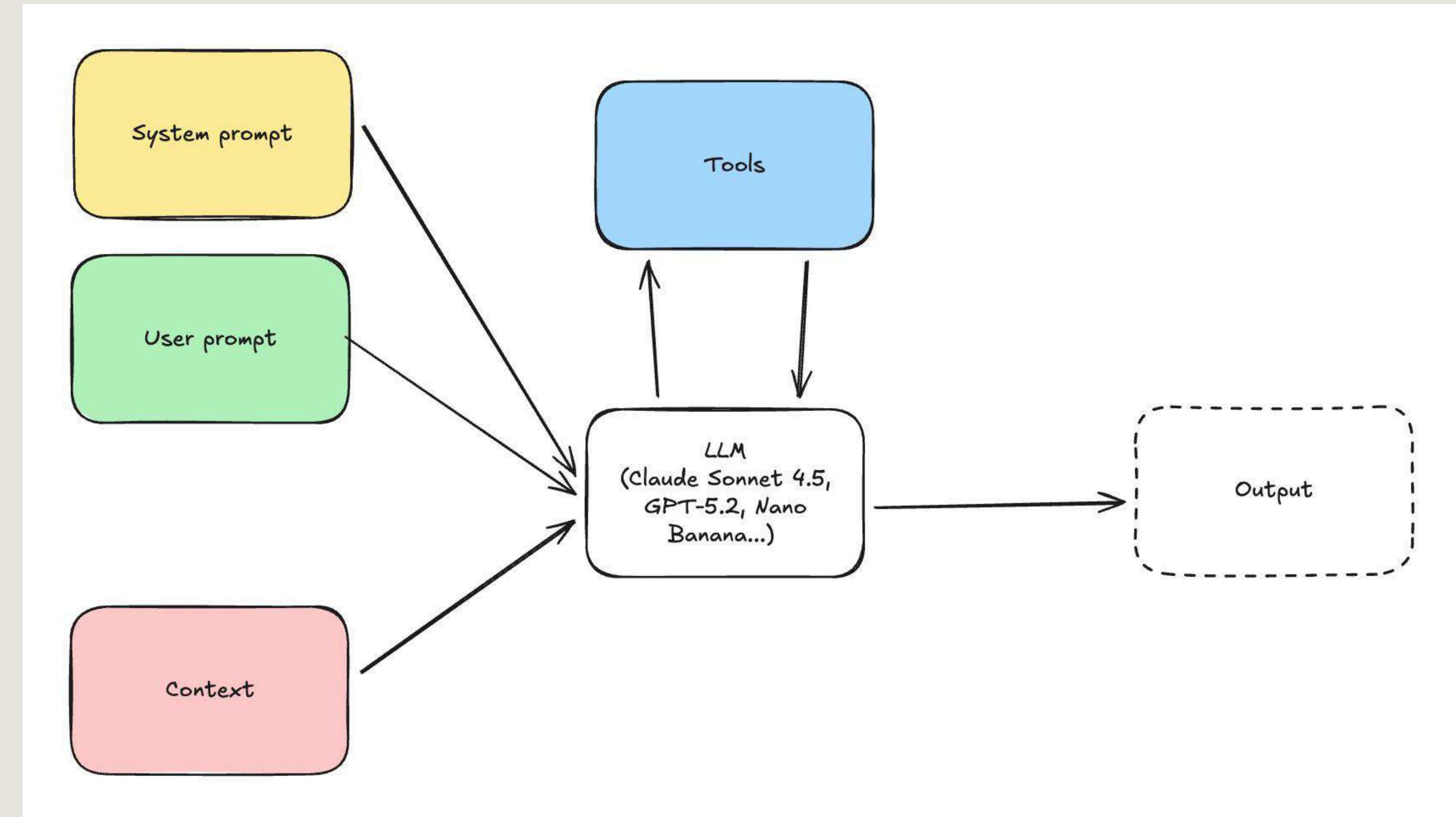


TOC IN TWO IMAGES: THE HOW



LET'S MODEL THE SYSTEM

GOAL: CONSISTENT HIGH QUALITY OUTPUTS

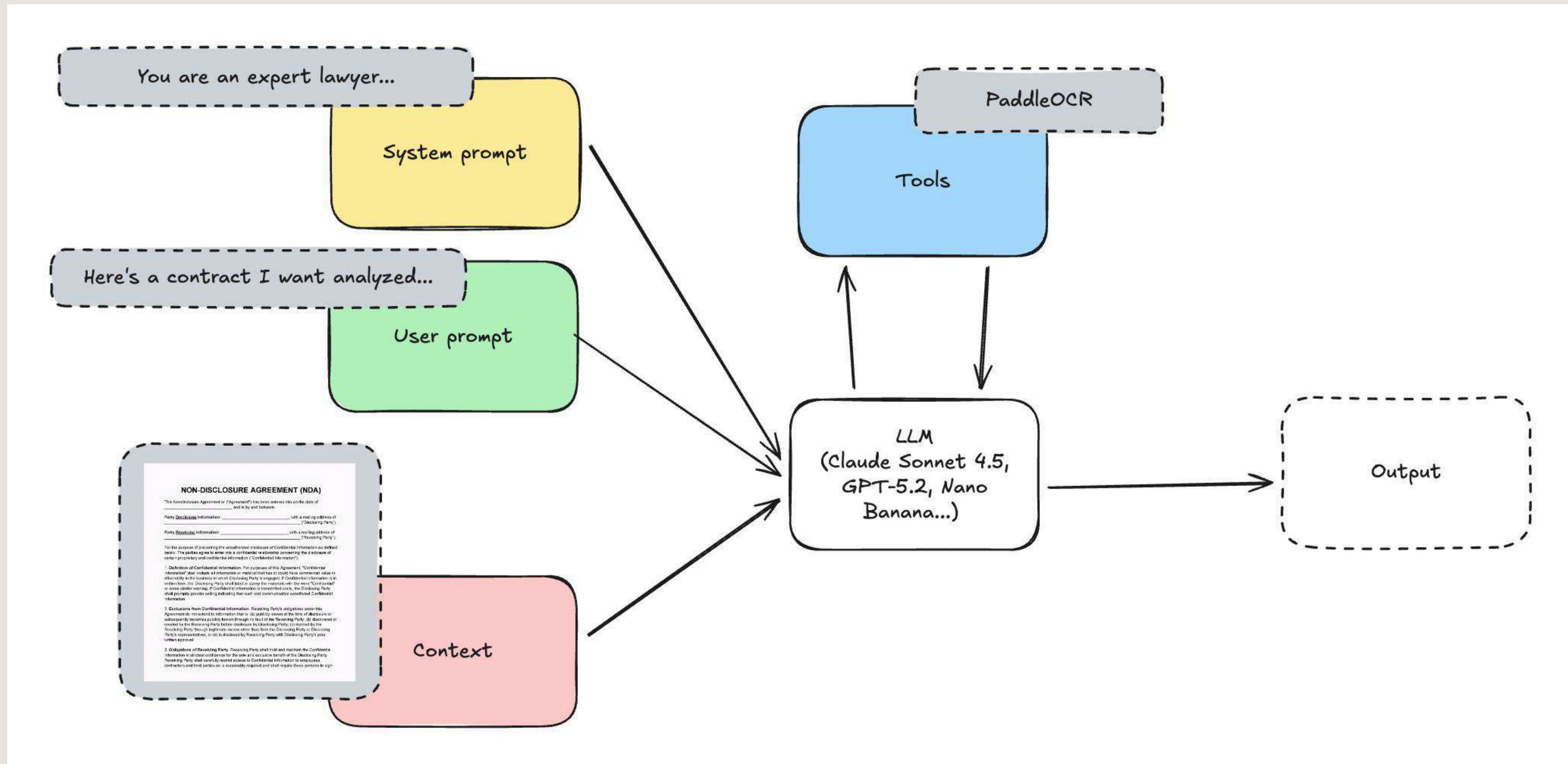


[HTTPS://ANDYBROMBERG.COM/FIELD-GUIDE-CONTEXT-ENGINEERING](https://andybromberg.com/field-guide-context-engineering)

[HTTPS://ANDYBROMBERG.COM/CONSTRAINTS-AI](https://andybromberg.com/constraints-ai)

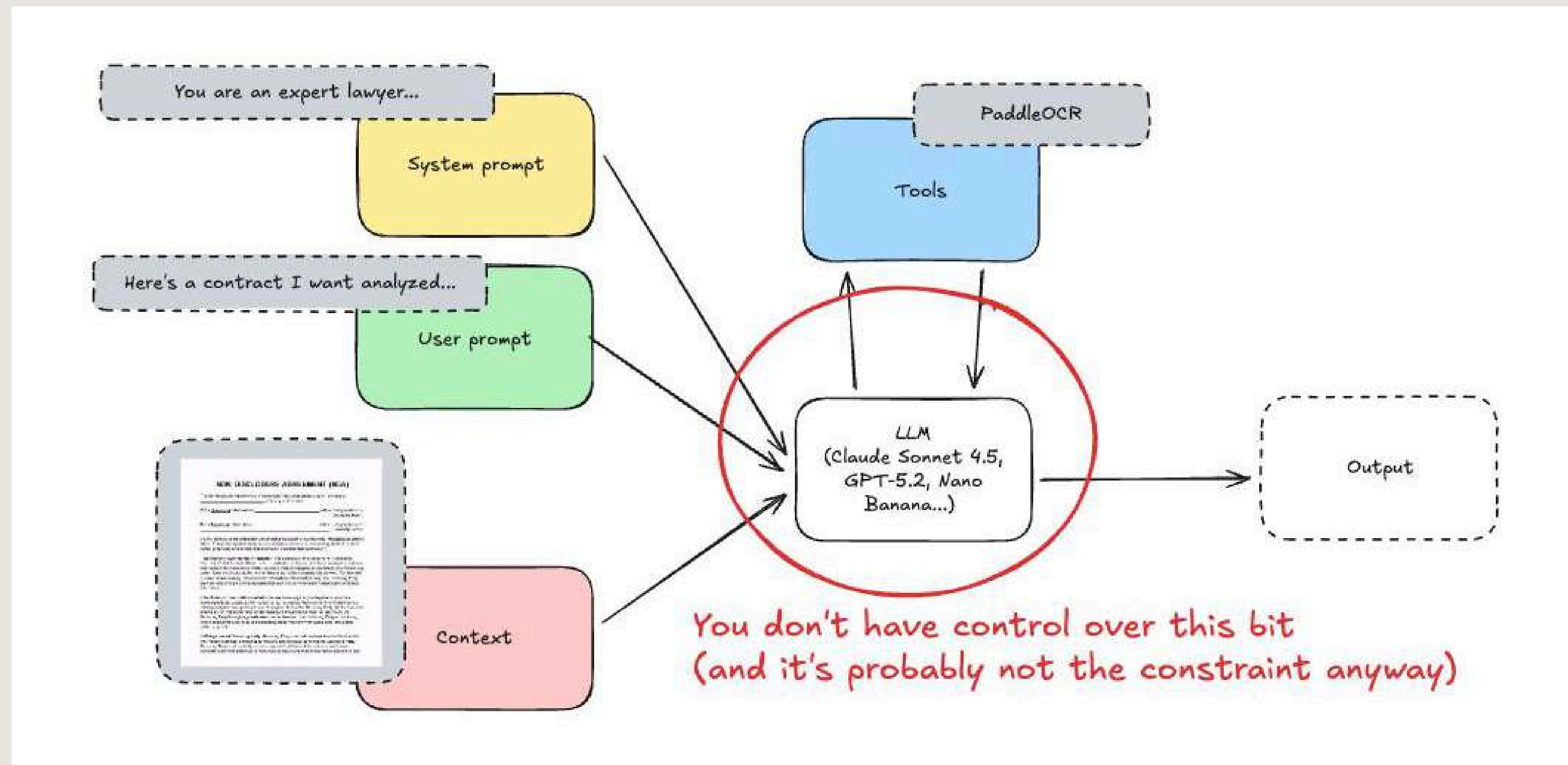
LET'S MODEL THE SYSTEM

GOAL: CONSISTENT HIGH QUALITY OUTPUTS



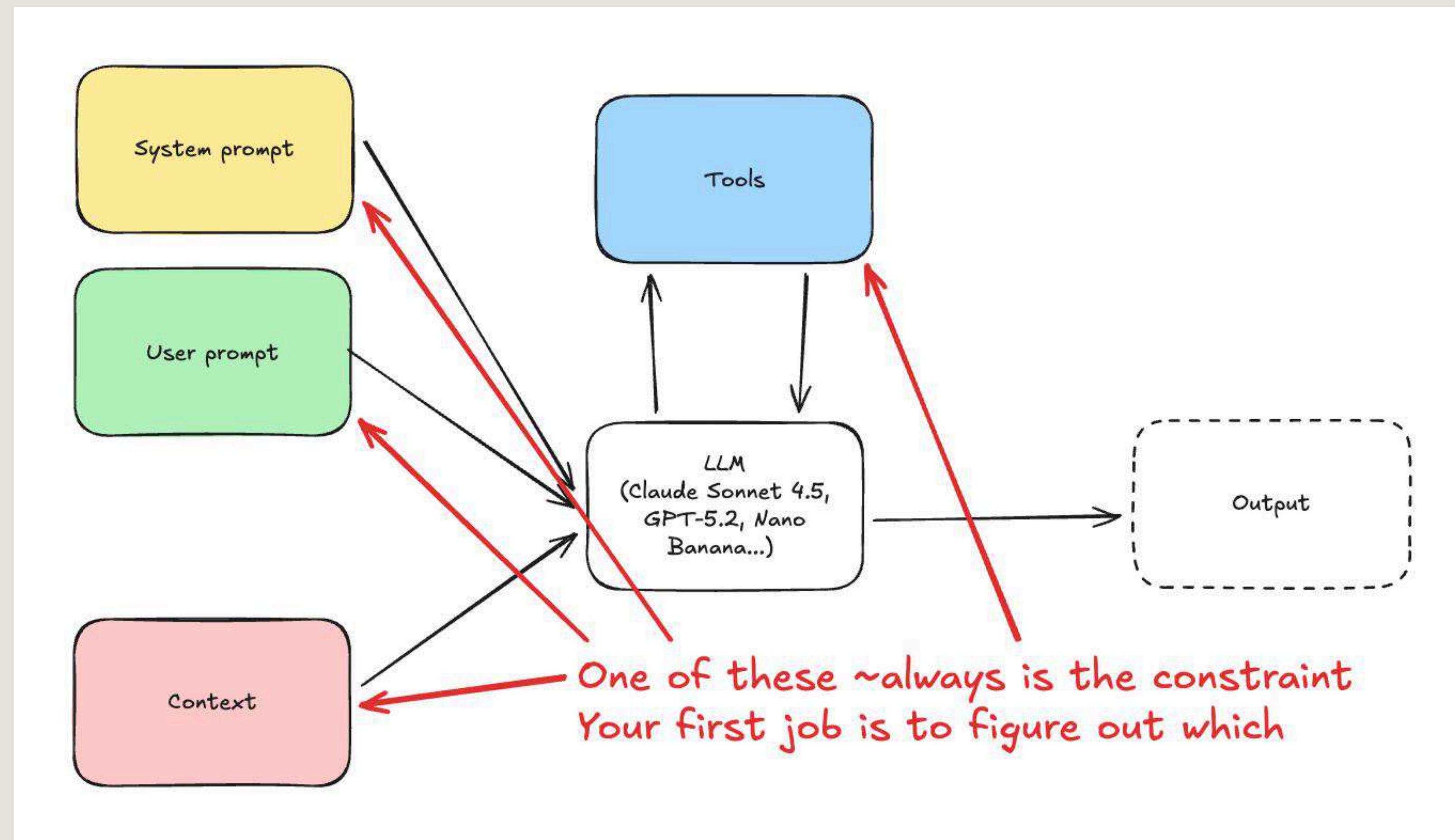
LET'S MODEL THE SYSTEM

GOAL: CONSISTENT HIGH QUALITY OUTPUTS

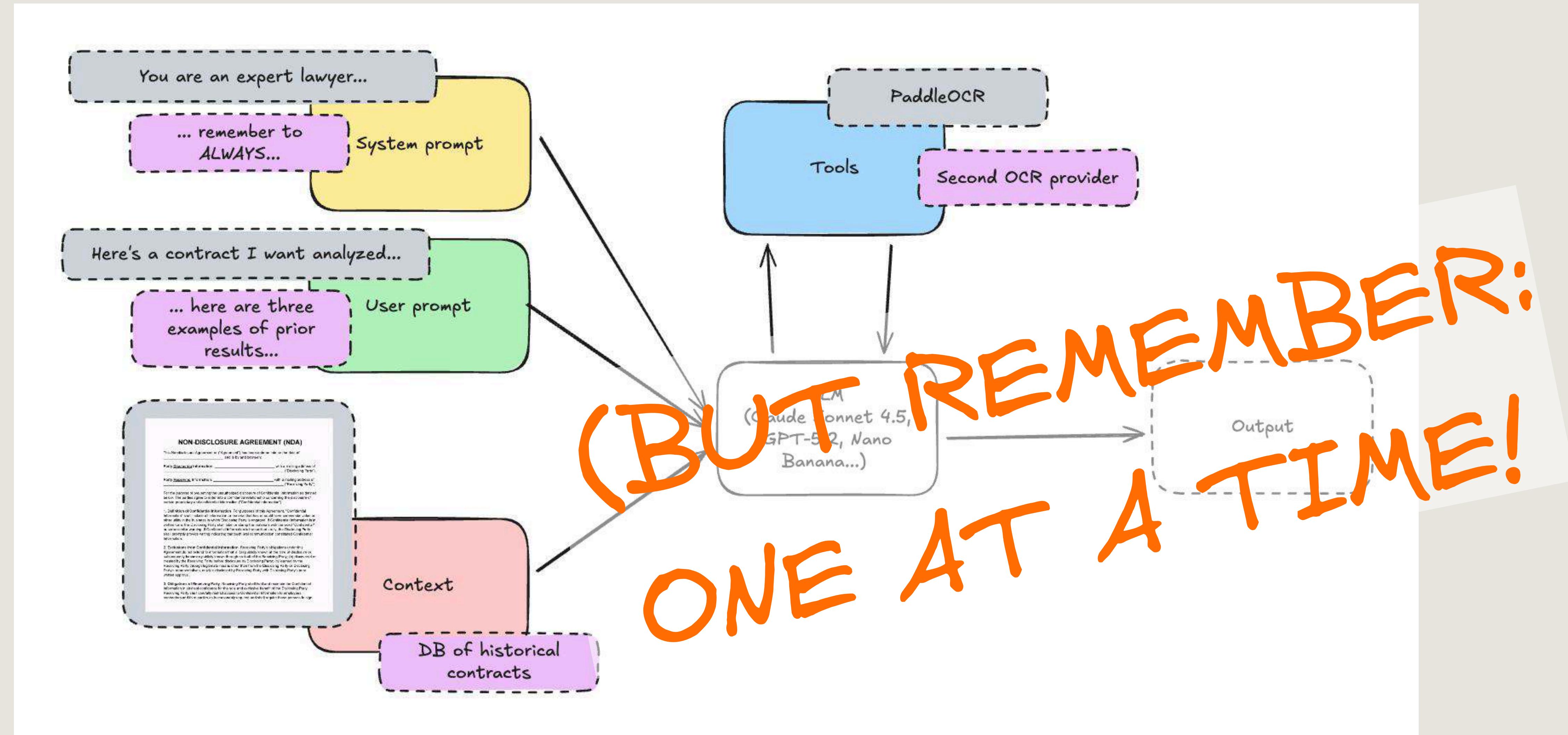


LET'S MODEL THE SYSTEM

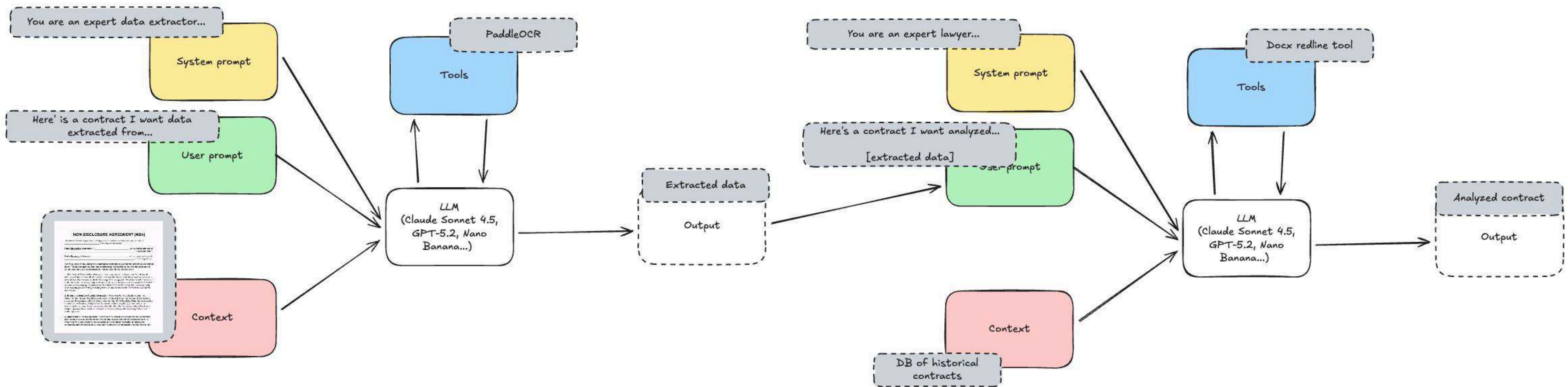
GOAL: CONSISTENT HIGH QUALITY OUTPUTS



WHAT MIGHT CHANGES LOOK LIKE?



SOMETIMES YOU NEED TO CHANGE YOUR VIEW OF THE SYSTEM

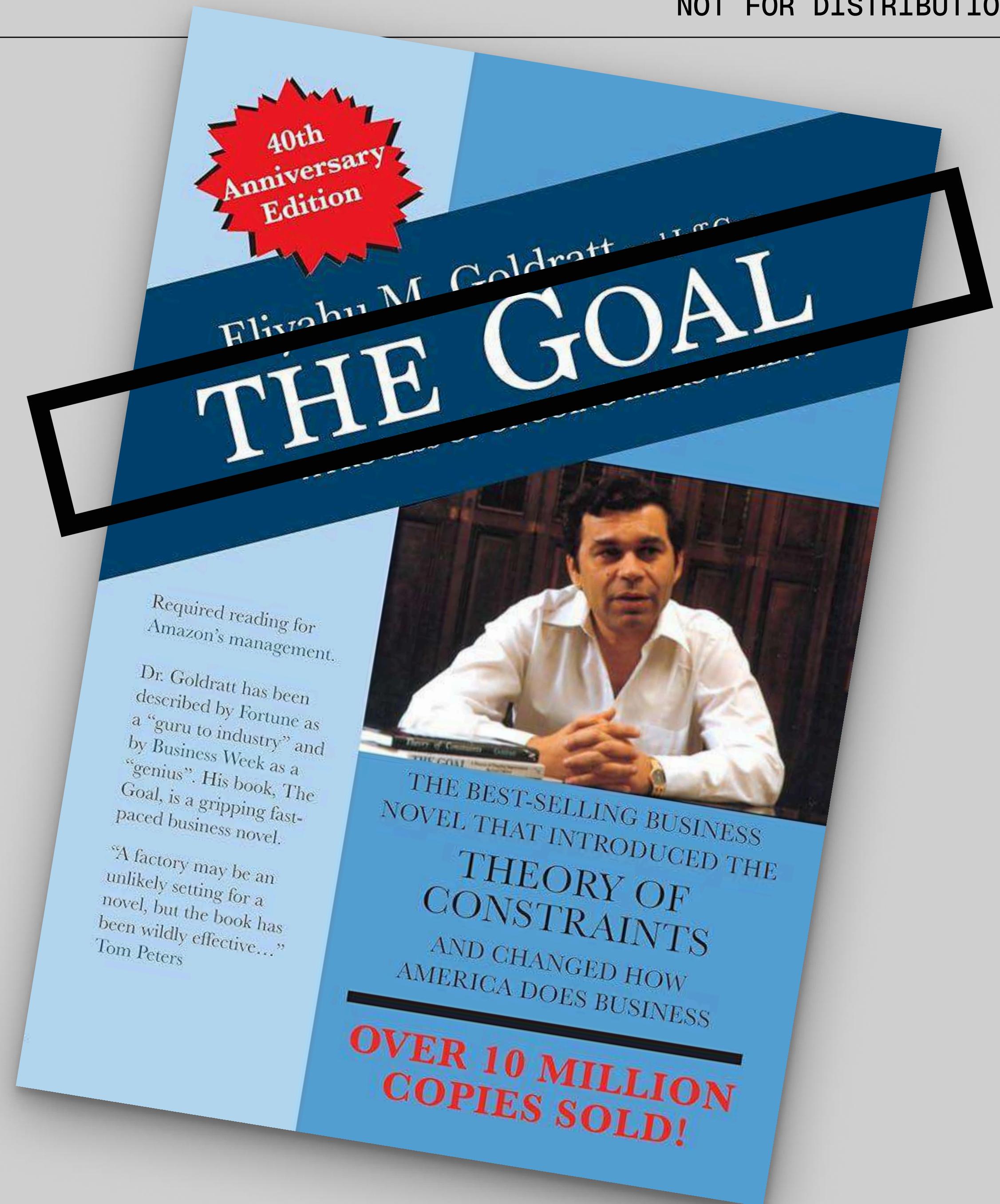


**HOW DO YOU KNOW WHAT TO CHANGE
AND IF IT HELPED?**

YOU NEED TO DEFINE ...

“YOU CANNOT UNDERSTAND THE MEANING OF PRODUCTIVITY UNLESS YOU KNOW WHAT THE GOAL IS.

UNTIL THEN, YOU'RE JUST PLAYING A LOT OF GAMES WITH NUMBERS AND WORDS.”



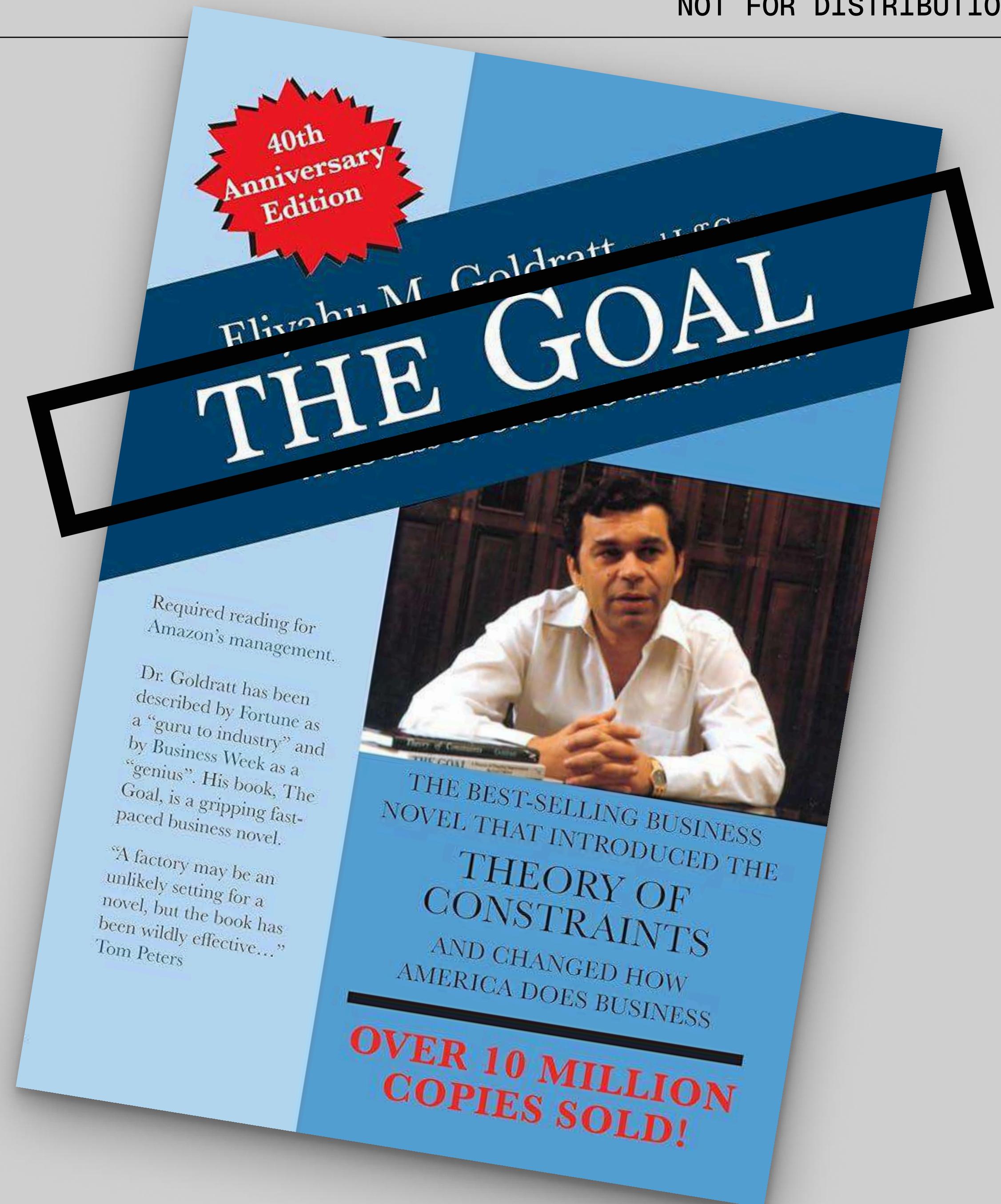
YOU NEED TO DEFINE ...

GOOD RESULTS NEED IMPROVEMENT

IMPROVEMENT NEEDS EVALS

EVALS NEED A GOAL

A GOAL NEEDS A DEFINITION



WE SAID OUR GOAL WAS:

CONSISTENT HIGH QUALITY OUTPUTS

BUT HOW DO YOU
MEASURE THAT?

HOW DO YOU MEASURE YOUR GOAL?

1. "PROPER" EVALS

(MAY YOU BE BLESSED
WITH A PRODUCT THAT
PERMITS THESE)

2. VIBE EVALS

PROPER EVALS

TL;DR:

- BUILD AND UPDATE AN “EVAL SET”
- RUN AGAINST IT CONTINUALLY
- UPDATE ACCORDINGLY
- ADD ONLINE/LIVE EVALUATION AS A BONUS

BUT: SCORING IS HARD FOR MANY USE CASES

(WHY DO YOU THINK MERCOR,
SCALE AI, SURGE AI ARE AS
BIG AS THEY ARE?)

braintrust

Galileo

POST <https://api.openai.com/v1/evals>

harbor

LangSmith

EXAMPLE: DOCUMENT EXTRACTION

PHOTOGRAPHY CONTRACT

EMPLOYMENT CONTRACT

This Employment Agreement ("Agreement") is made and entered into, _____, 20_____, by and between:

EMPLOYER: [Employer's name and address] _____
Employee: [Employee's full name and address] _____

The Employee is employed by the Employer as an employee at the Employer's place of business, and the Employer agrees to employ the Employee for the term of this Agreement.

The Employer has the right to terminate the employment relationship at any time, with or without cause, and the Employee has the right to resign at any time.

I. EMPLOYEE'S DUTIES: The Employee agrees to perform his/her job duties in accordance with the Employer's reasonable direction. The Employee will use his/her best efforts to perform his/her job duties in a professional manner and to the best of his/her ability. The Employee will not engage in any conduct that would be considered a violation of the law or regulations, or any conduct that would be considered unprofessional or unethical. The Employee agrees to follow all reasonable instructions, guidelines, policies, and procedures set forth by the Employer. The Employer may change its business address, hours of work, and other terms and conditions of employment at any time.

II. COMPENSATION: The Employee will be paid a gross salary of \$_____. The Employer will deduct taxes, insurance, and other expenses from the Employee's paychecks.

III. EMPLOYMENT PERIOD: The Employee agrees to serve the Employer for _____ years.

A. TERM: If the Employee resigns or is terminated, he/she will be entitled to _____ days notice. If the Employee resigns, he/she will be entitled to _____ days notice. If the Employee is terminated, he/she will be entitled to _____ days notice. If the Employee is terminated, he/she will be entitled to _____ days notice.

B. TERMINATION: The Employer reserves the right to terminate the Agreement for cause, _____ days notice. The Employer reserves the right to terminate the Agreement for convenience, _____ days notice. The Employer reserves the right to terminate the Agreement for non-compliance with the terms of the Agreement, _____ days notice.

C. PAYMENT: The Employer will pay the Employee _____ per hour for _____ hours worked per week. The Employer will pay the Employee _____ per month for _____ months worked per year. The Employer will pay the Employee _____ per year for _____ years worked.

D. BENEFITS: The Employer will provide the Employee with _____ benefits.

E. OBLIGATIONS: The Employee agrees to keep the Employer's information confidential.

F. GOVERNING LAW: This Agreement shall be governed by the laws of the state of _____.

G. DISPUTE RESOLUTION: Any dispute arising under this Agreement shall be resolved through arbitration in accordance with the rules of the American Arbitration Association. The arbitration award shall be final and binding.

H. ENTIRE AGREEMENT: This Agreement contains the entire understanding between the parties and supersedes all prior agreements between the parties.

I. AMENDMENTS: Any changes to this Agreement must be in writing and signed by both parties.

J. SEVERABILITY: If any provision of this Agreement is held invalid or unenforceable, the remaining provisions shall remain valid and enforceable.

K. WAIVER: The Employer waives the right to require the Employee to waive his/her rights under this Agreement.

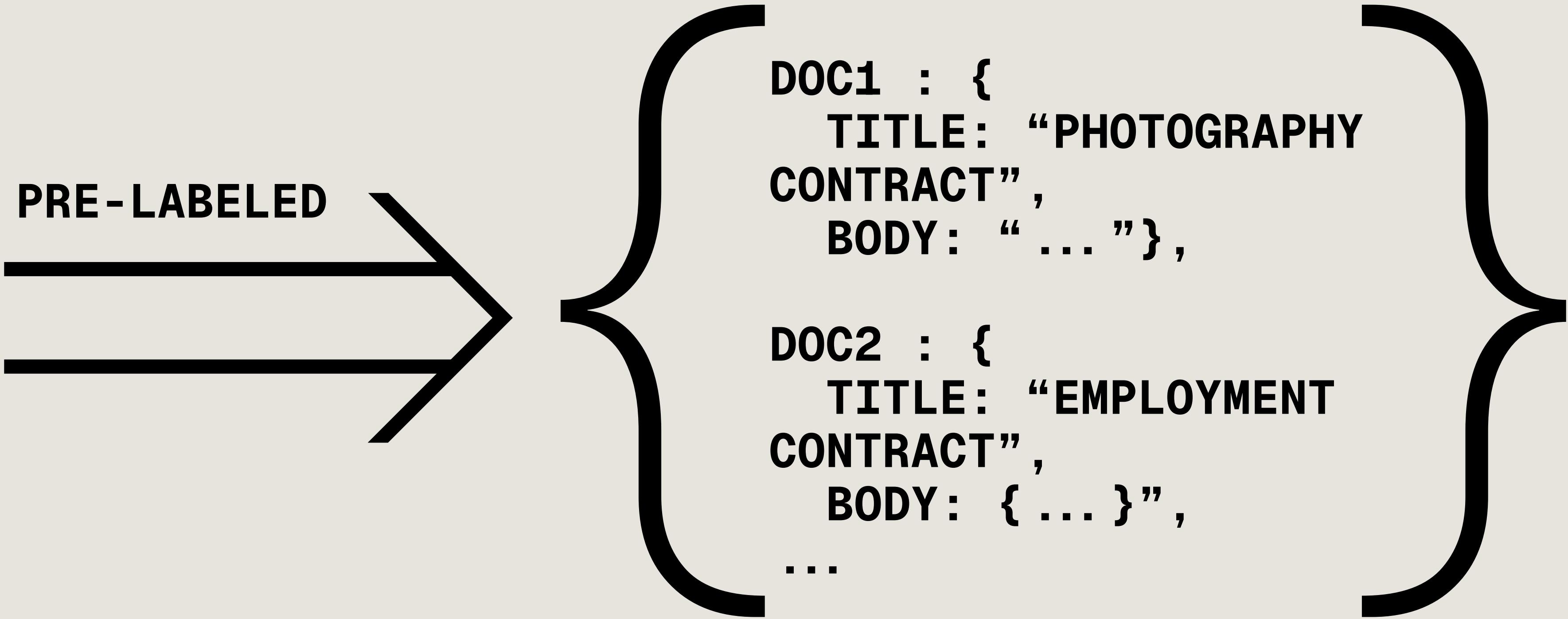
L. NOTICES: All notices shall be in writing and sent to the Employer at _____ and to the Employee at _____.

M. SIGNATURES: The Employer and the Employee have read and understood this Agreement and agree to be bound by its terms.

EMPLOYER'S SIGNATURE: _____ Date: _____

EMPLOYEE'S SIGNATURE: _____ Date: _____

PRE-LABELED



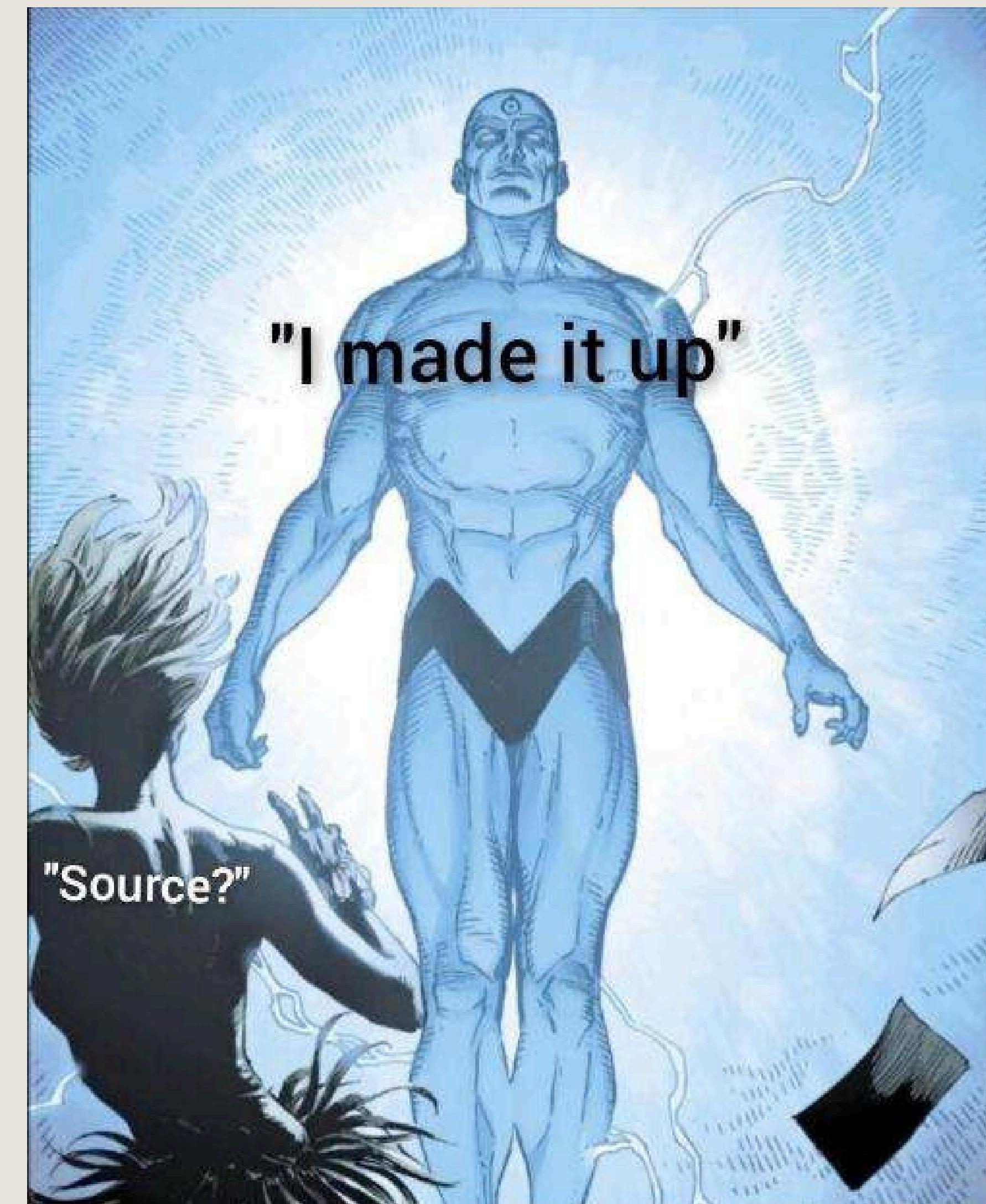
SCORE = SIMILARITY({PRE-LABELED}, {ACTUAL OUTPUT})

GOAL = MAXIMIZE(SCORE)

VIBE EVALS

TL;DR:

- JUST, UH, LOOK AT THE RESULTS AND ASK YOURSELF HOW GOOD THEY ARE

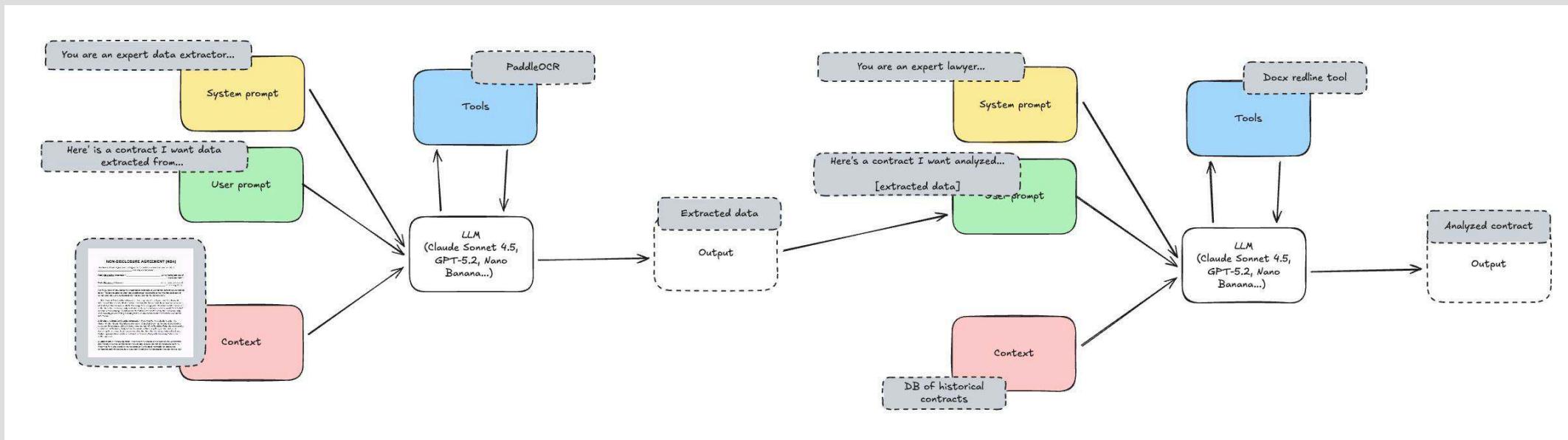


EVAL OPTIONS

- 1 PROPER EVAL: PRE-LABELED RESULTS THAT SHOULD MATCH EXACTLY
- 2 PROPER EVAL: PRE-LABELED RESULTS THAT YOU CAN COMPUTE SIMILARITY TO
- 3 PROPER EVAL: LLM AS A JUDGE OF THE RESULTS
- 4 VIBE EVAL: HOW DOES IT FEEL

IN SUMMARY: THE IMPROVEMENT LOOP

STEP 1: DIAGRAM OUT YOUR SYSTEM

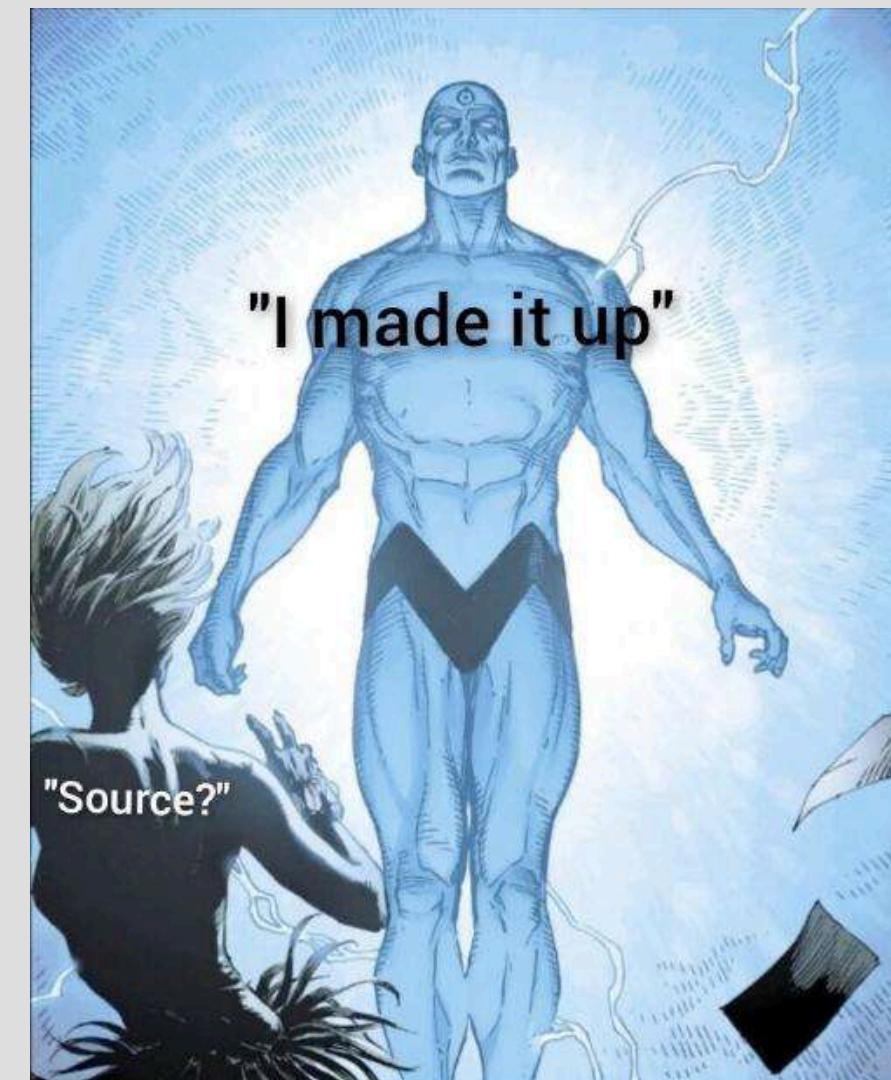


STEP 2: DEFINE YOUR GOAL & EVAL PROCESS FOR SAID GOAL "PROPER" ... OR ... VIBE EVALS?

Scorer types

Braintrust offers three types of scorers:

- **Autoevals**: Pre-built, battle-tested scorers for common tasks like factuality checking, semantic similarity, and format validation. Start here for standard evaluation needs.
- **LLM-as-a-judge**: Use a language model to evaluate outputs based on natural language criteria. Best for subjective judgments like tone, helpfulness, or creativity that are difficult to encode in code.
- **Custom code**: Write custom evaluation logic in TypeScript or Python. Best when you have specific rules, patterns, or calculations to implement. Custom code scorers can evaluate either the final output or the entire execution trace for multi-step workflows.

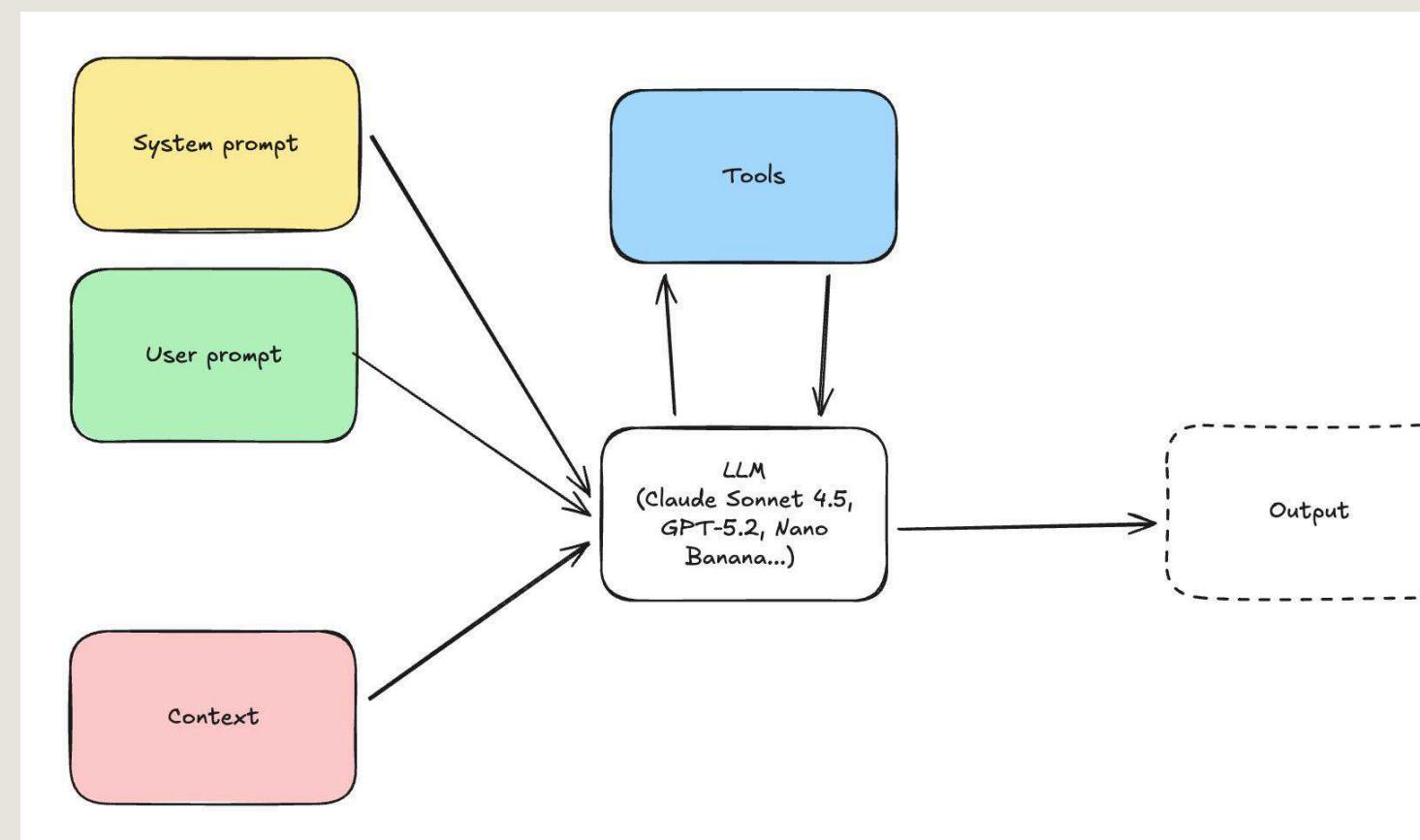


IN SUMMARY: THE IMPROVEMENT LOOP

STEP 1: DIAGRAM OUT YOUR SYSTEM

STEP 2: DEFINE YOUR GOAL & EVAL PROCESS FOR SAID GOAL

STEP 3: ASSESS WHERE YOU THINK THE CONSTRAINT IS



SYSTEM PROMPT, USER PROMPT,
CONTEXT, OR TOOLS
... OR THE SYSTEM VIEW ITSELF

**STEP 4: USE TOC: EXPLOIT CONSTRAINT, SUBORDINATE EVERYTHING ELSE;
ELEVATE CONSTRAINT ... AND MEASURE EVALS EACH TIME**

IN SUMMARY: THE IMPROVEMENT LOOP

STEP 1: DIAGRAM OUT YOUR SYSTEM

STEP 2: DEFINE YOUR GOAL & EVAL PROCESS FOR SAID GOAL

STEP 3: ASSESS WHERE YOU THINK THE CONSTRAINT IS

**STEP 4: USE TOC: EXPLOIT CONSTRAINT, SUBORDINATE EVERYTHING ELSE;
ELEVATE CONSTRAINT ... AND MEASURE EVALS EACH TIME**

STEP 5: GOTO STEP 3

IN SUMMARY: THE IMPROVEMENT LOOP

STEP 1: DIAGRAM OUT YOUR SYSTEM

STEP 2: DEFINE YOUR GOAL & EVAL PROCESS FOR SAID GOAL

STEP 3: ASSESS WHERE YOU THINK THE CONSTRAINT IS

STEP 4: USE TOC: EXPLOIT CONSTRAINT, SUBORDINATE EVERYTHING ELSE;
ELEVATE CONSTRAINT ... AND MEASURE EVALS EACH TIME

STEP 5: GOTO STEP 3

THIS SOUNDS BORING AND
PROGRAMMATIC AND OBVIOUS...
BUT... IT WORKS

(AND MOST PEOPLE DO NOT APPROACH IT SO RIGOROUSLY)

IN SUMMARY: THE IMPROVEMENT LOOP

STEP 1: DIAGRAM OUT YOUR SYSTEM

STEP 2: DEFINE YOUR GOAL & EVAL PROCESS FOR SAID GOAL

STEP 3: ASSESS WHERE YOU THINK THE CONSTRAINT IS

STEP 4: USE TOC: EXPLOIT CONSTRAINT, SUBORDINATE EVERYTHING ELSE;
ELEVATE CONSTRAINT ... AND MEASURE EVALS EACH TIME

STEP 5: GOTO STEP 3

THIS SOUNDS BORING AND
PROGRAMMATIC AND OBVIOUS...
BUT... IT WORKS

BONUS: IT ACTUALLY WORKS FOR MOST
ANY PROBLEM IN LIFE OR BUSINESS

THREE PRACTICAL EXAMPLES

1 LIGHTWORK KNOWLEDGE BASE

2 INTERFACEØ

3 [REDACTED]



EVAL:

- HOLDBACK OF A SET OF GOOD EMAILS (WITH THE CLIENT INQUIRY) AND SEGMENTS OF REPORTS (WITH THE RAW DATA)
- RUN LATEST AND GREATEST AGAINST THE INPUTS
- USE AN LLM AS A JUDGE TO COMPARE THE HOLDBACK WITH THE OUTPUTS

TOC APPLICATION:

- ORIGINAL CONSTRAINT: SYSTEM PROMPT
- ONGOING CONSTRAINT: CONTEXT (MORE AND MORE AND MORE)

Lightwork Knowledge Base

[Ask](#) [Manage](#)

Ask

Assessment context (optional)

Search assessments...

Response type

Draft email

Global contexts

251111 Collated Reports ×

Lightwork Newsletters ×

Evaluation Standards ×

Search web as well Show debug

Question

a client is doing a big renovation and asked about what recessed lighting options would be good

3 contexts selected

Subject: Recessed Lighting Recommendations for Your Renovation

Hi [Client Name],

It's great to hear about the renovation. Since you are starting with a clean slate, you have a unique opportunity to improve both your health and the aesthetic of the home.

When selecting recessed lighting, we prioritize three main health metrics to avoid issues like eyestrain, headaches,

1. Flicker-Free Technology

Most standard LEDs flicker because of how they convert AC electricity to DC. Even if it's not visible, this can cause

interface0

- EVAL:**
- VIBES - THE REQUESTS ARE TOO VARIED FOR A PROPER EVAL

interface0



How can I help you today?

+ Knowledge

You can /use prompts or @mention knowledge & tools

GPT-5.2

0

File icon

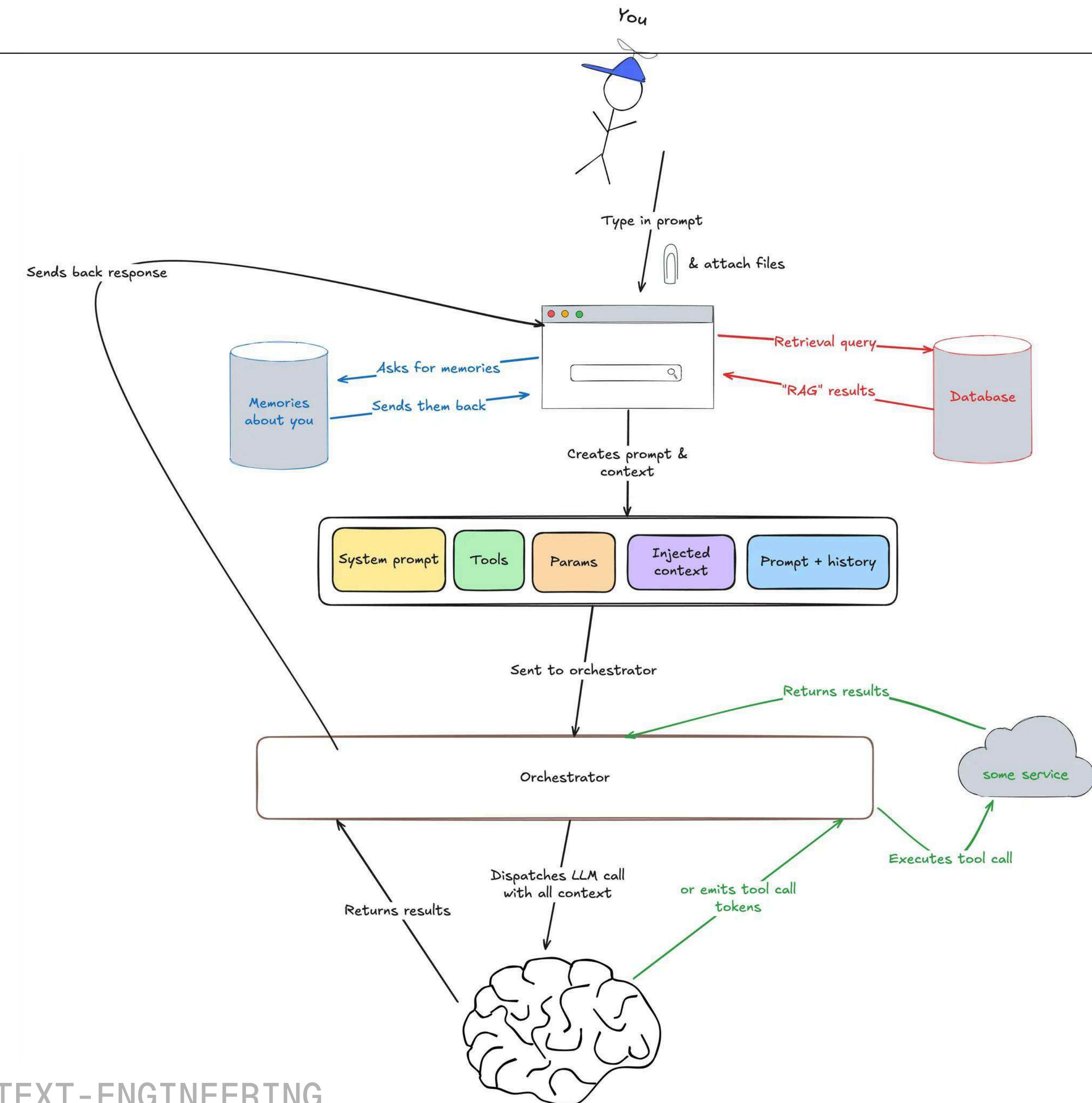
Microphone icon

Star icon

Up arrow icon

interface0

- EVAL:**
- VIBES - THE REQUESTS ARE TOO VARIED FOR A PROPER EVAL
- TOC APPLICATION:**
- ORIGINAL CONSTRAINT: SYSTEM PROMPT
 - SECOND CONSTRAINT: USER PROMPTS → HELPED WITH TEMPLATES
 - ONGOING CONSTRAINTS: ALTERNATING BETWEEN TOOLS AND CONTEXT

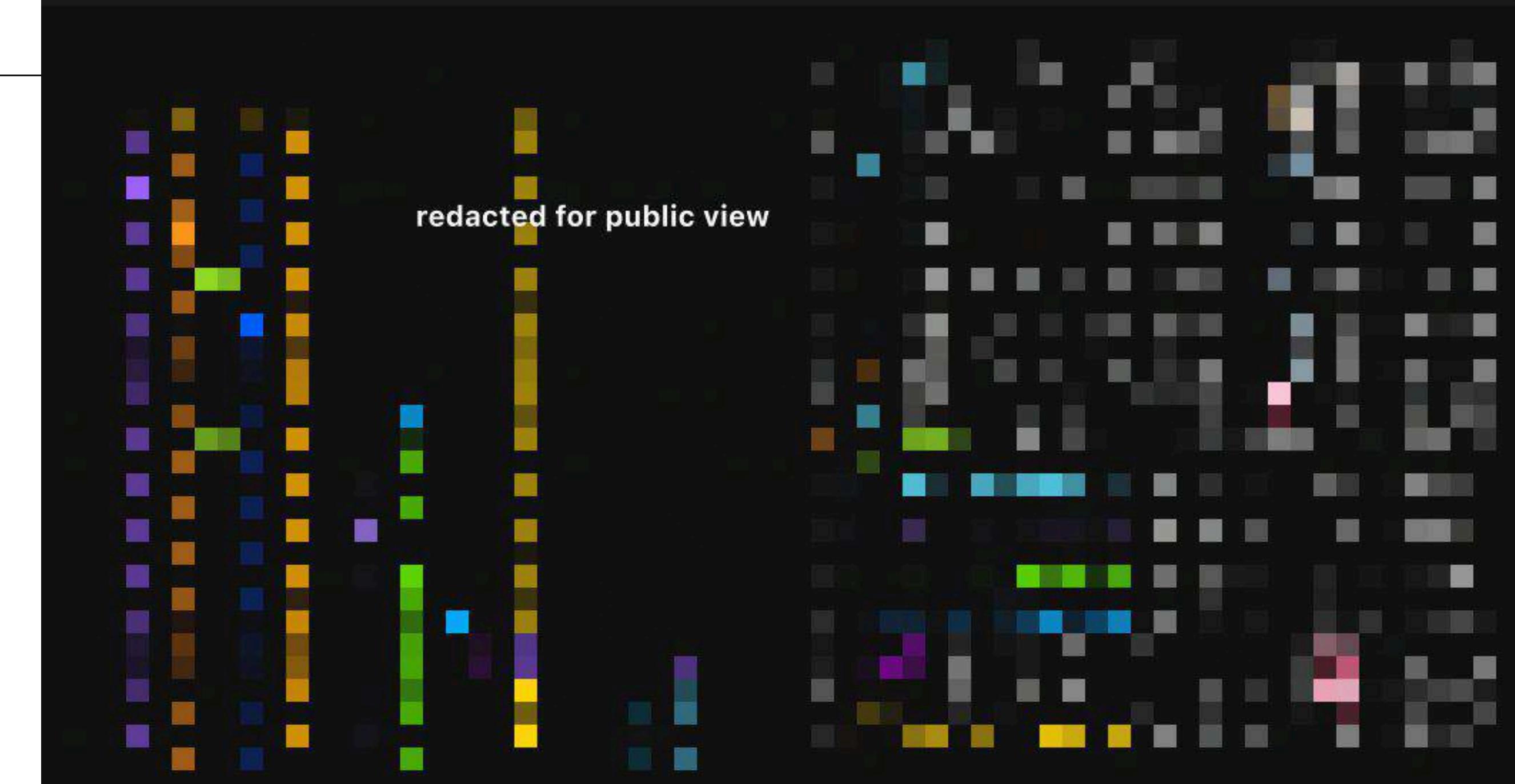


[HTTPS://ANDYBROMBERG.COM/FIELD-GUIDE-CONTEXT-ENGINEERING](https://andybromberg.com/field-guide-context-engineering)

[REDACTED]

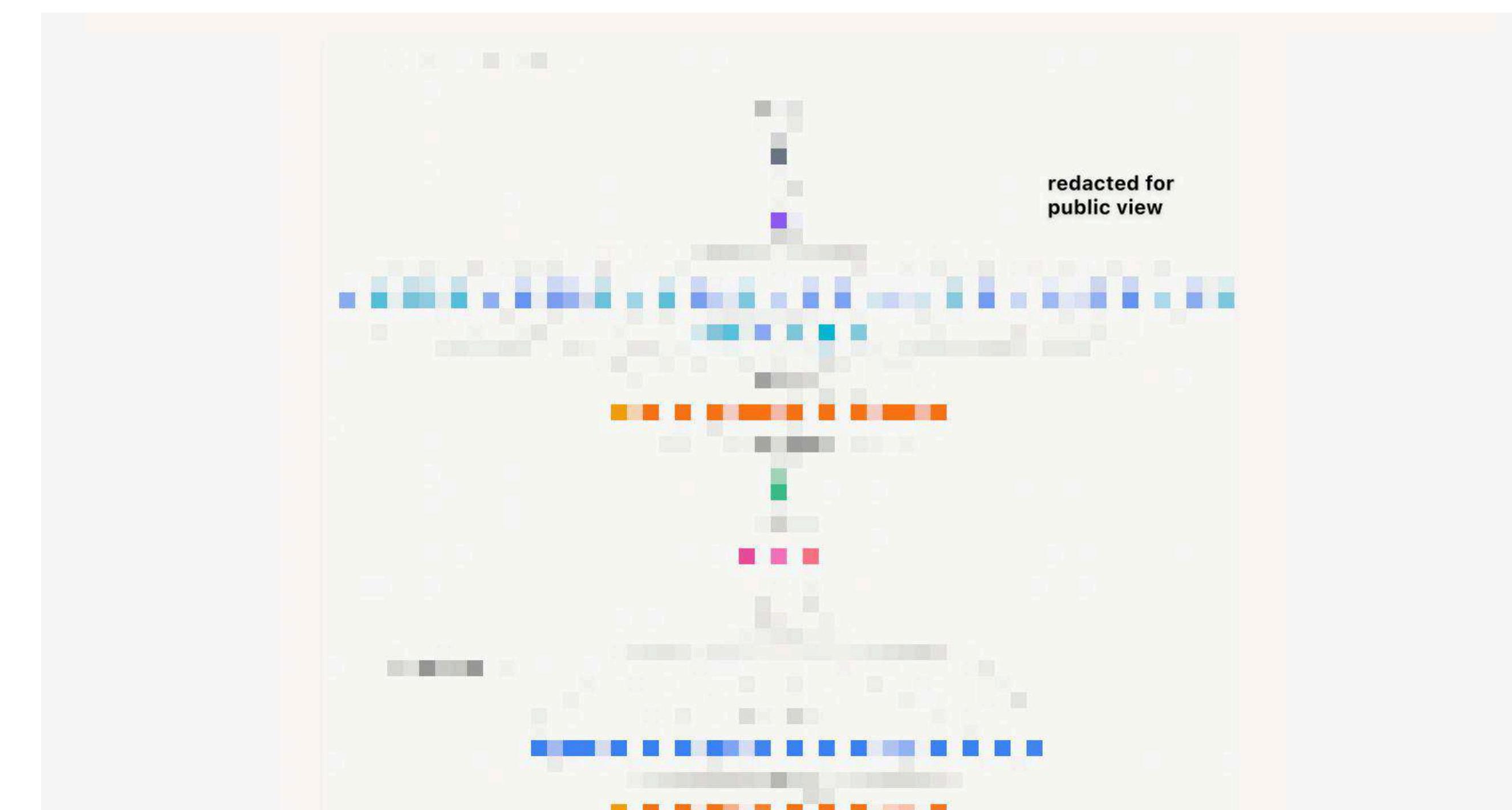
EVAL:

- EXPERT FEEDBACK ON THE RESULTS (SOME STRUCTURED, SOME UNSTRUCTURED)
- USER FEEDBACK ON THE RESULTS (UNSTRUCTURED)



TOC APPLICATION:

- ORIGINAL CONSTRAINT: SYSTEM PROMPTS FOR SPECIFIC SUBAGENTS
- THEN: TOOLS
- NOW BACK TO CONTEXT & SYSTEM PROMPTS



WHAT WE COVERED

-
- 1 THE ISSUE WITH LLMS
 - 2 THE THEORY OF CONSTRAINTS
 - 3 EVALS & HOW TO IMPROVE
 - 4 SOME PRACTICAL EXAMPLES
-

THE STATE OF PLAY TODAY:

THE LLM IS ~NEVER THE CONSTRAINT

SO IF WE BUY THE THEORY OF CONSTRAINTS ...

THAT MEANS HERE ARE THINGS IN YOUR CONTROL TO GET
YOUR PRODUCT TO THE QUALITY & CONSISTENCY YOU WANT

IT IS YOUR JOB TO MAP THE SYSTEM, BUILD EVALS,
FIND THE CURRENT CONSTRAINT, AND FIX IT

RINSE & REPEAT. MAGIC IS IN YOUR HANDS!

PLEASE REACH OUT!

ANDY@ANDYBROMBERG.COM

I'D LOVE TO HEAR FROM YOU :)

P.S.: HTTPS://TINKERING.CLUB

TOOLS I CAN'T LIVE WITHOUT*



Trigger.dev

Build and deploy fully-managed AI agents and workflows

Trigger.dev is the platform for building AI workflows in TypeScript. Long-running tasks with retries, queues, observability, and elastic scaling.

[Start building now >](#)

13.5k Open source

braintrust

Turn insights into improvements with [Loop](#)

Iterate, eval, ship

Braintrust is the AI observability platform for building quality AI products

[Get started](#) [Request a demo](#)

Claude API Docs

Agent SDK > Overview

Agent SDK overview

Build production AI agents with Claude Code as a library

TextGrad

Open in Colab License MIT Nature Paper arXiv 2406.07496 docs passing python 3.9 pypi v0.1.8 platform noarch
conda-forge v0.1.4

TextGrad: Automatic "Differentiation" via Text

An autograd engine -- for textual gradients!

EVERY



Midjourney/Every illustration.

Chain of Thought

Compound Engineering: How Every Codes With Agents

* PROBABLY OUT OF DATE BY FEBRUARY 6 2025 AT THE PACE THINGS ARE GOING

BUILDING GOOD LLM PRODUCTS WITH THE THEORY OF CONSTRAINTS

STANFORD CS 224G

FEBRUARY 5 2026

ANDY BROMBERG
ANDY@ANDYBROMBERG.COM
TINKERING: ANDYBROMBERG.COM · INVESTING: AMITY.VC