



# The New Source Code

Jan 15, 2026

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

# Outline

1. **The demo is a lie**
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

McKinsey  
& Company

**87%**

of LLM-based projects never  
make it to production

**Gartner**

**40%**

will be canceled by 2027 due to  
... inadequate risk controls

# THE HILL

BUSINESS

## Air Canada must pay refund promised by AI chatbot, tribunal rules

BY NICK ROBERTSON - 02/18/24 11:05 PM ET

Democrats irked over 'pro-choice' GOP candidates

Canada must pay a Vancouver man a partial refund ticket that was promised by the Canadian tribunal ruled Wednesday, in a landmark case for the use of AI in business.

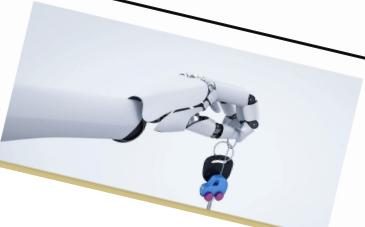
Inc.

### A Chevrolet Dealership Used ChatGPT for Customer Service and Learned That AI Isn't Always on Your Side

This is why you need to narrow your AI's scope before deploying it.

BY BEN GREGORY, STAFF REPORTER BENLUCA@THEHILL.COM

DEC 18, 2023



# BBC

## Why Google's 'woke' AI problem won't be an easy fix

28 February 2024

Zoe Kleinman  
Technology editor

Share Save

**Most Popular**

- 1 Pilot flying Helene in rescue missions in threatening with a
- 2 Democrats sub Neta to tell

**Psychiatrist**

In the last few days, Google's artificial intelligence (AI) tool Gemini has had to face



**MOST POPULAR**

**CASE REPORT**

Black Hairy Tongue: A Rare Adverse Effect of Olanzapine

A young woman recently diagnosed with a manic episode developed black discoloration on the dorsum of her tongue 3 weeks after starting olanzapine.

*Prim Care Companion CNS Disord*  
2024;26(5):24cr03757  
Preeti Datal and others

**CASE REPORT**

Age-Friendly Health Systems: The Medical and Mental Health Connection

# Outline

1. The demo is a lie
2. **What is a prompt**
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

Traditional software

**Same input → Same output**

AI applications

**Same input → Different output**

It's not just outputs.

**Inputs changed too.**

Text. Images. Audio. Video. Documents.

Context-dependent. Conversation-aware.

What prompt actually encode

**Business logic**

**Company knowledge**

**Operational intelligence**

**Safety constraints**

**User instructions**

...

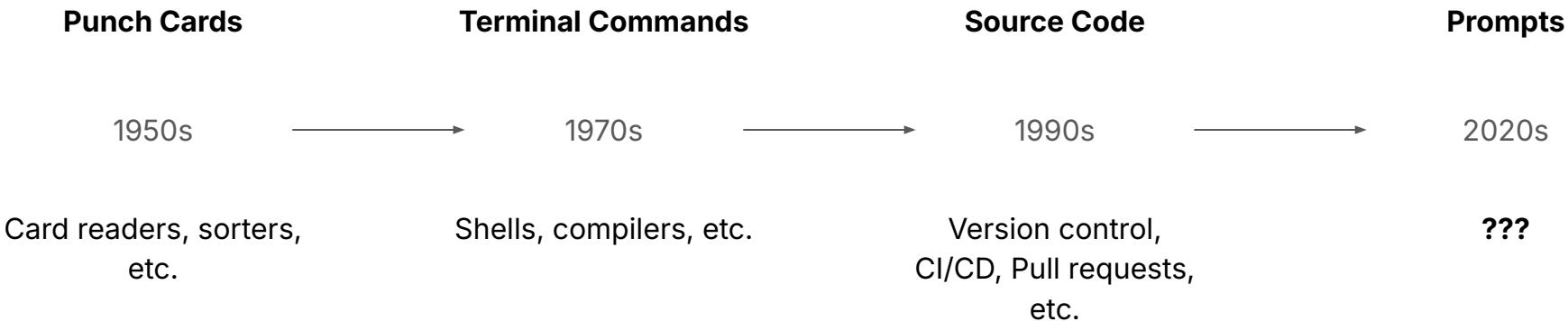
Prompts are not configuration.

**Prompts are executable logic.**

# Outline

1. The demo is a lie
2. What is a prompt
3. **Artifacts of the past of computing**
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

Every computing paradigm  
**creates new artifacts.**



New artifacts need  
**new infrastructure.**

What we built for code over decades

**Version control**

**Testing & CI/CD**

**Environment management**

**Monitoring & Observability**

...

**Prompts need all of this.**

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. **The AI development lifecycle**
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

# Iterate

# Evaluate

# Monitor

# Deploy

continuous cycle



# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. **Iterate**
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

**Iterate**

From guesswork to engineering

~~Copy. Paste. Pray.~~

**Structured experiments**

**Version history with diffs**

**Side-by-side prompting**

**Use-case driven development**

**DEMO**

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate**
  - c. Deploy
  - d. Monitor
5. Sum

Evaluate

From vibes to metrics

**Outputs vary.**

**Humans don't scale.**

**Simple metrics miss nuance.**

**You don't know "good" until you see it.**

Evals are not tests.

**Evals are feedback loops.**

The goal isn't to pass the eval.

**The goal is to learn and iterate faster.**

~~Write evals first.~~

**Look at the data first.**

You need a criteria to grade outputs.

**But grading outputs helps you define criteria.**

When vibes and data disagree,  
**trust the vibes.**

**DEMO**

LLM as a Judge

**Always binary. Pass or fail.**

**Never use a 1-10 scales.**

Nobody knows what 3.5 vs 7.2 means.

# 4-7

evals are enough for most applications

**Engineers build infrastructure.**

**Domain experts write rubrics.**

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. **Deploy**
  - d. Monitor
5. Sum

Deploy

From chaos to control

Frequency of prompt changes

≠

Frequency of code changes

But we still need

**Version control**

**Rollback**

**Audit trails**

**Environments**

...

**Engineers.**

**Domain experts.**

**DEMO**

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor**
5. Sum

**Monitor**

**Close the loop. Build the flywheel.**

**Real data**

**Datasets**

**Better prompts**

**Evals**

Data Flywheel



Run your evals  
**continuously on your prod data.**

**DEMO**

# Outline

1. The demo is a lie
2. What is a prompt
3. Artifacts of the past of computing
4. The AI development lifecycle
  - a. Iterate
  - b. Evaluate
  - c. Deploy
  - d. Monitor
5. Sum

# Iterate

# Evaluate

# Monitor

# Deploy

continuous cycle



Where do you start today

- 1. Build one use case**
- 2. Look at 20 traces**
- 3. Build a 20-example dataset**
- 4. Iterate and version your prompt**
- 5. Add basic monitoring**

Prompts are the new  
**business logic layer.**

Prompts are the new  
**new source code.**

# The New Source Code

arsh@adaline.ai