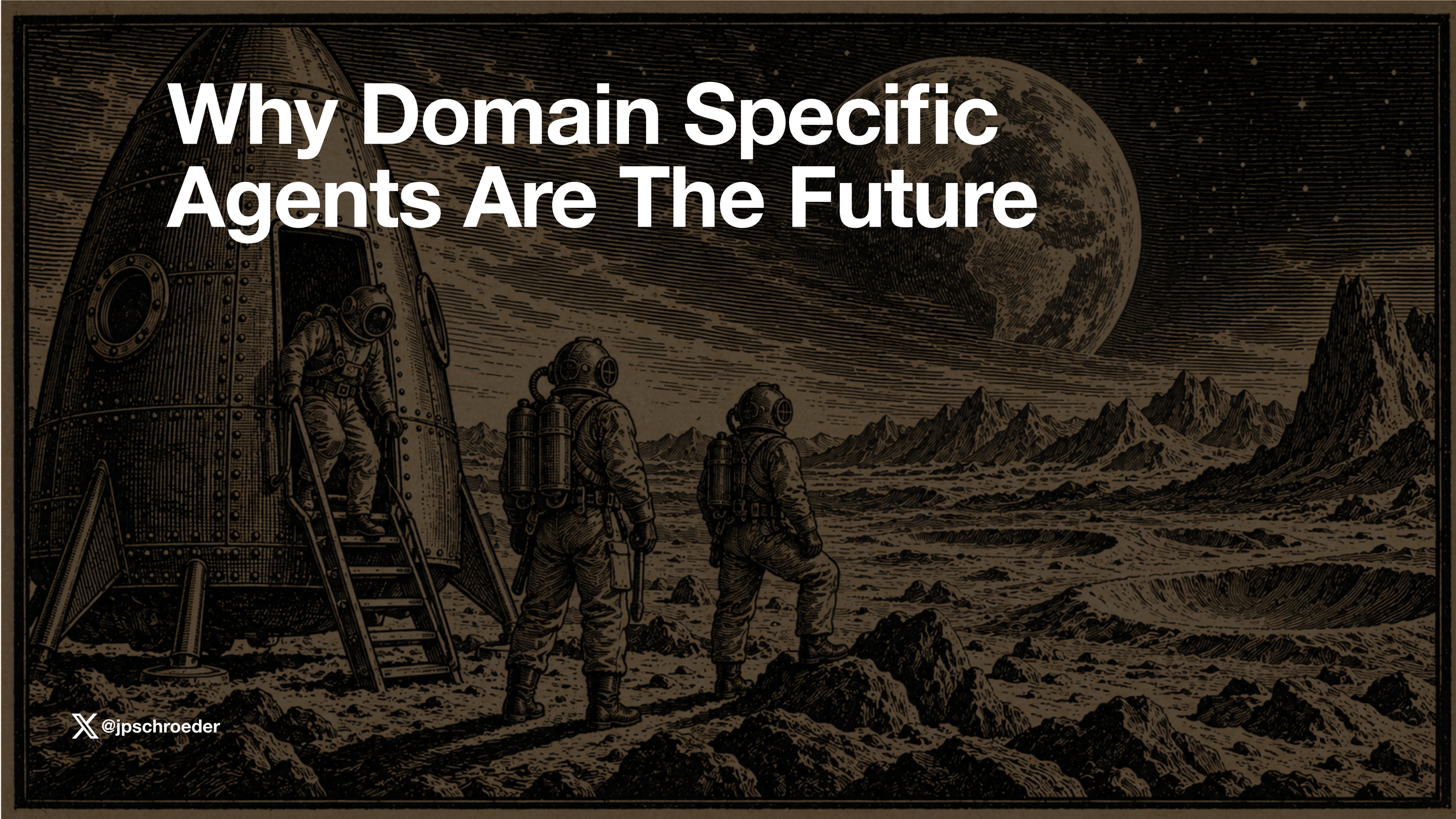
Why Domain Specific Agents Are The Future



@jpschroeder



Justin Schroeder

@jpschroeder

BRAID

StandardAgents

dmux

arrow JS

{...} jsonreader

ZOOOWN

FormKit

AutoAnimate

TEMPO

DRAG + DROP



SIP

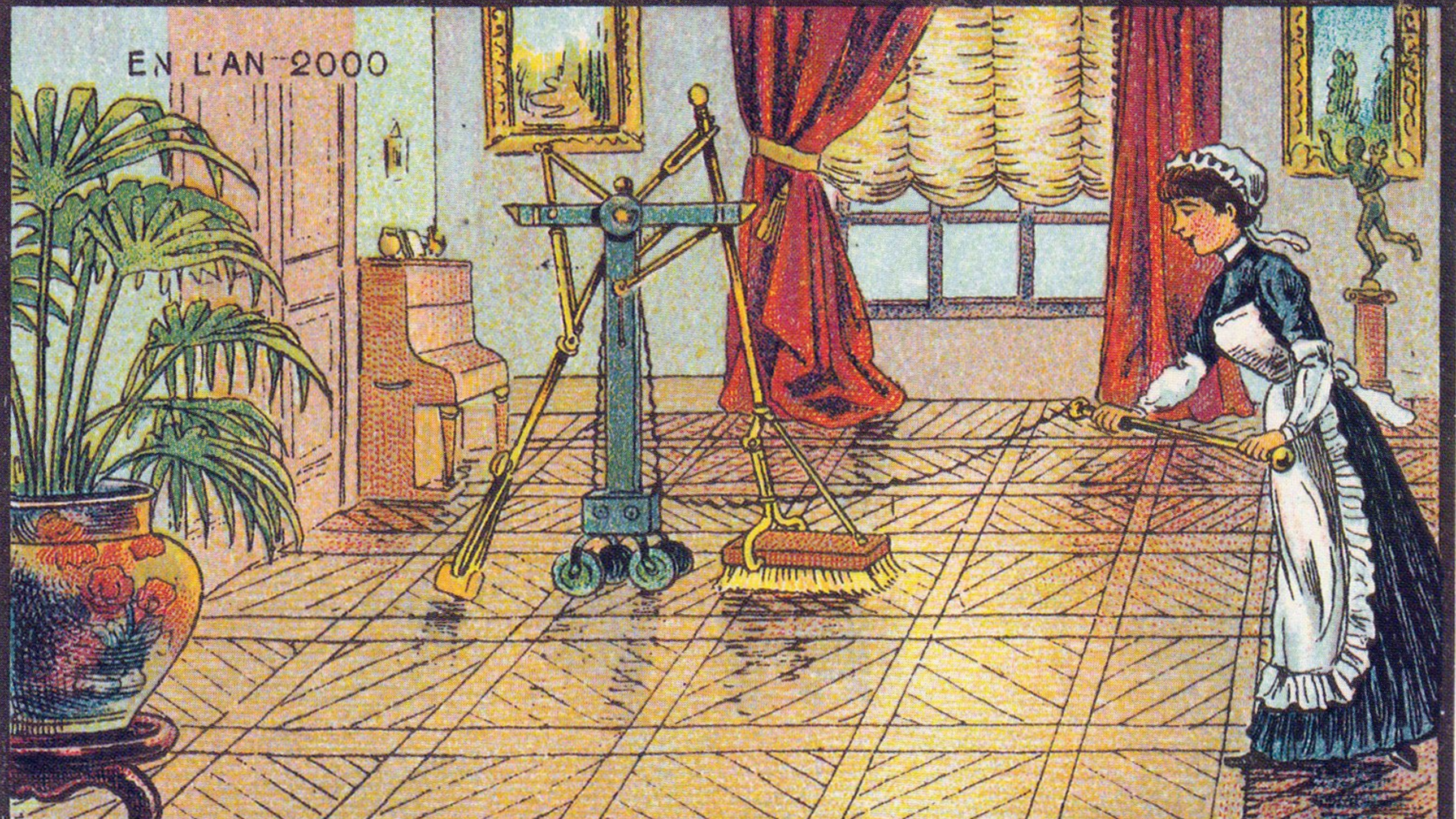
Small image processor

@jpschroeder

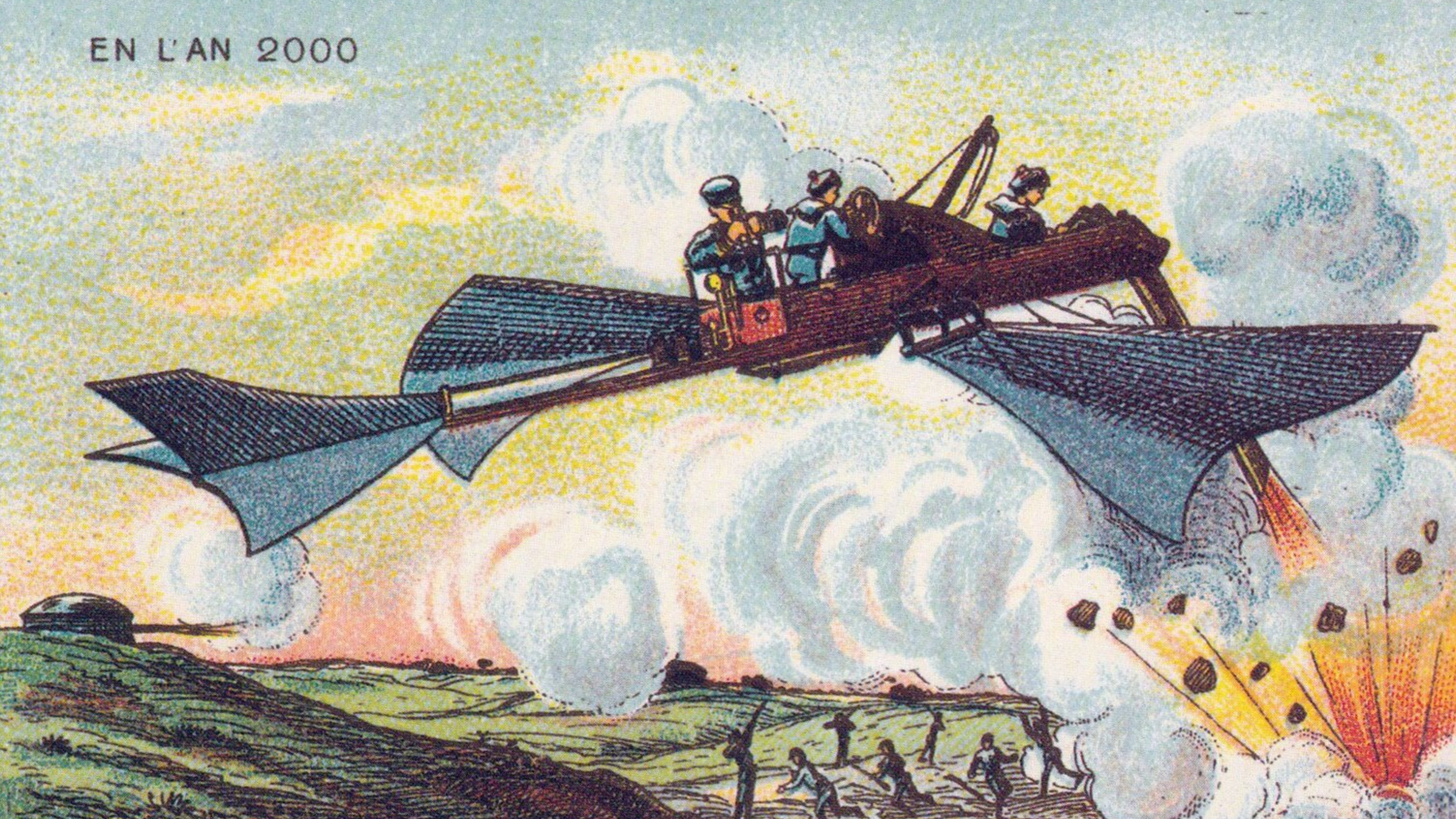
WARNING:

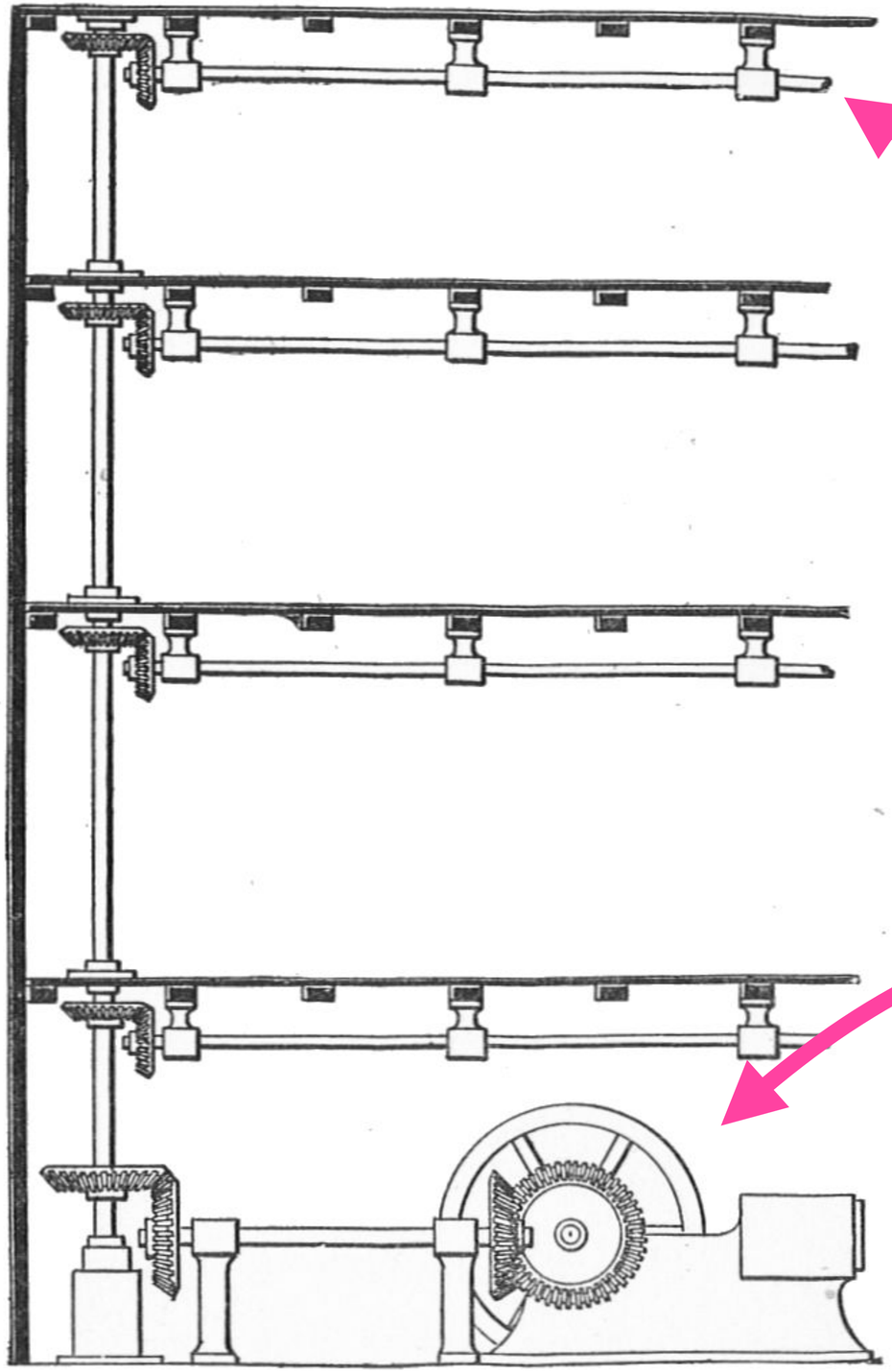
This talk contains forward looking statements.


EN L'AN 2000


EN L'AN 2000

EN L'AN 2000

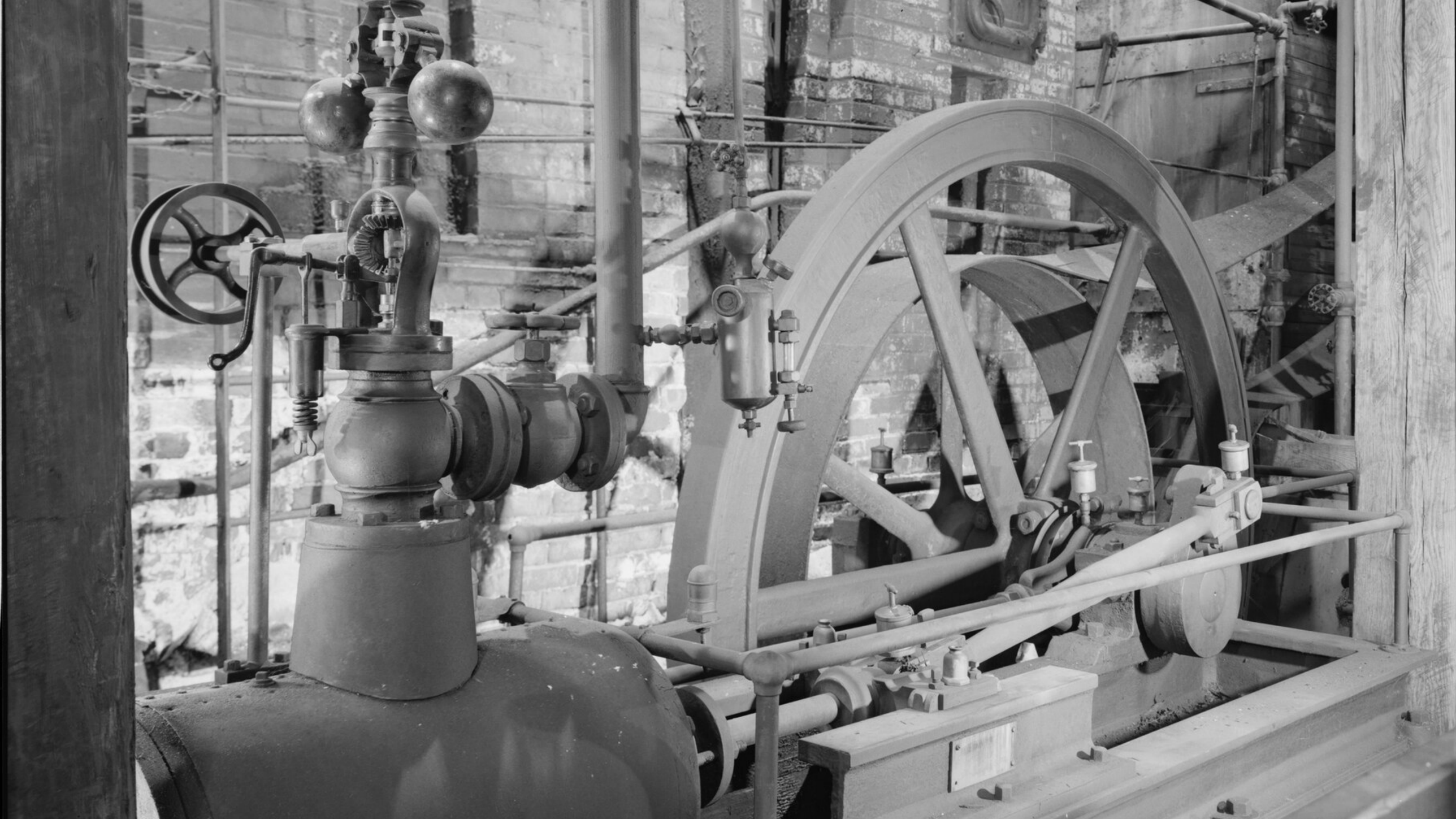




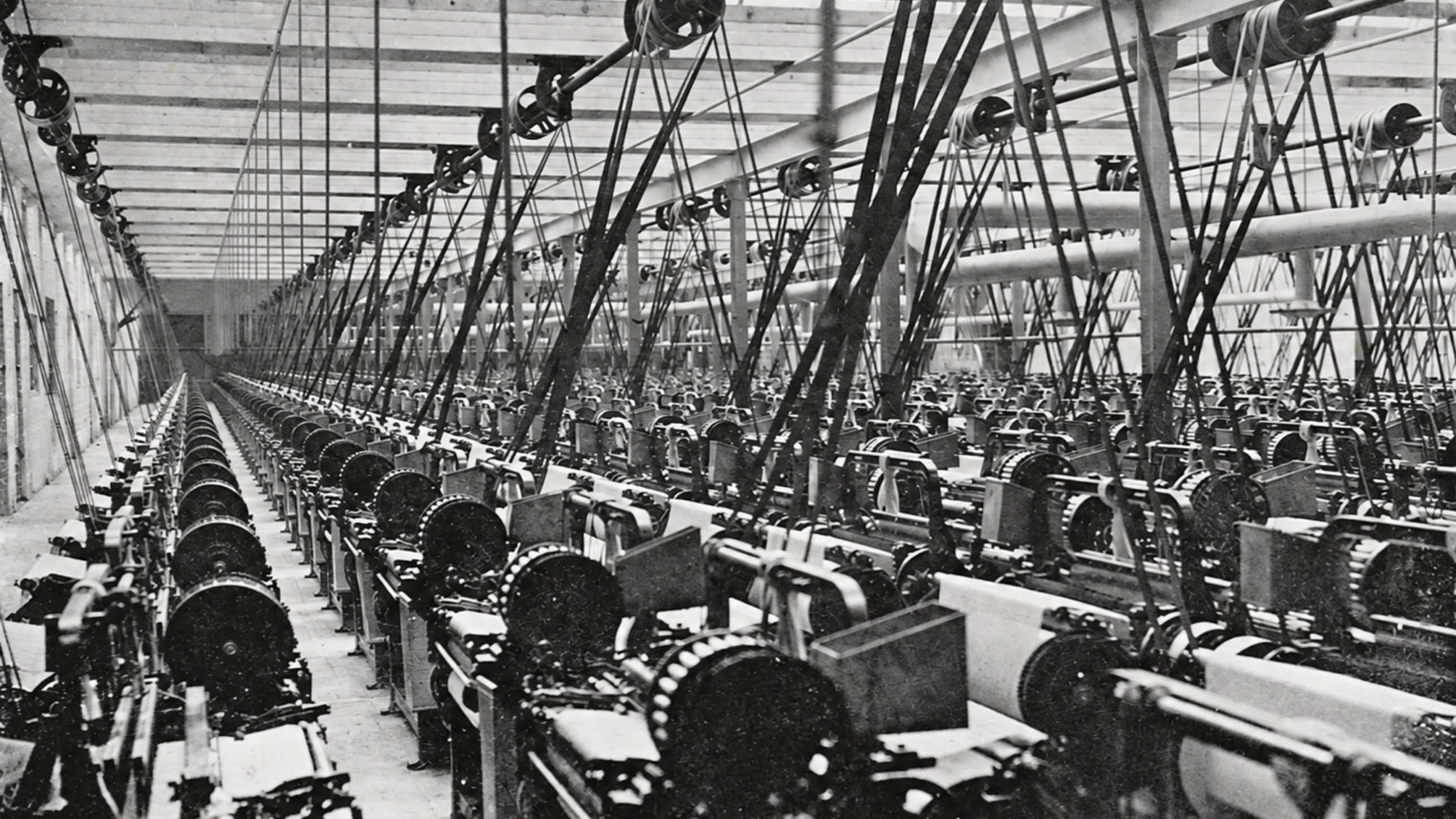
Rotating line shaft

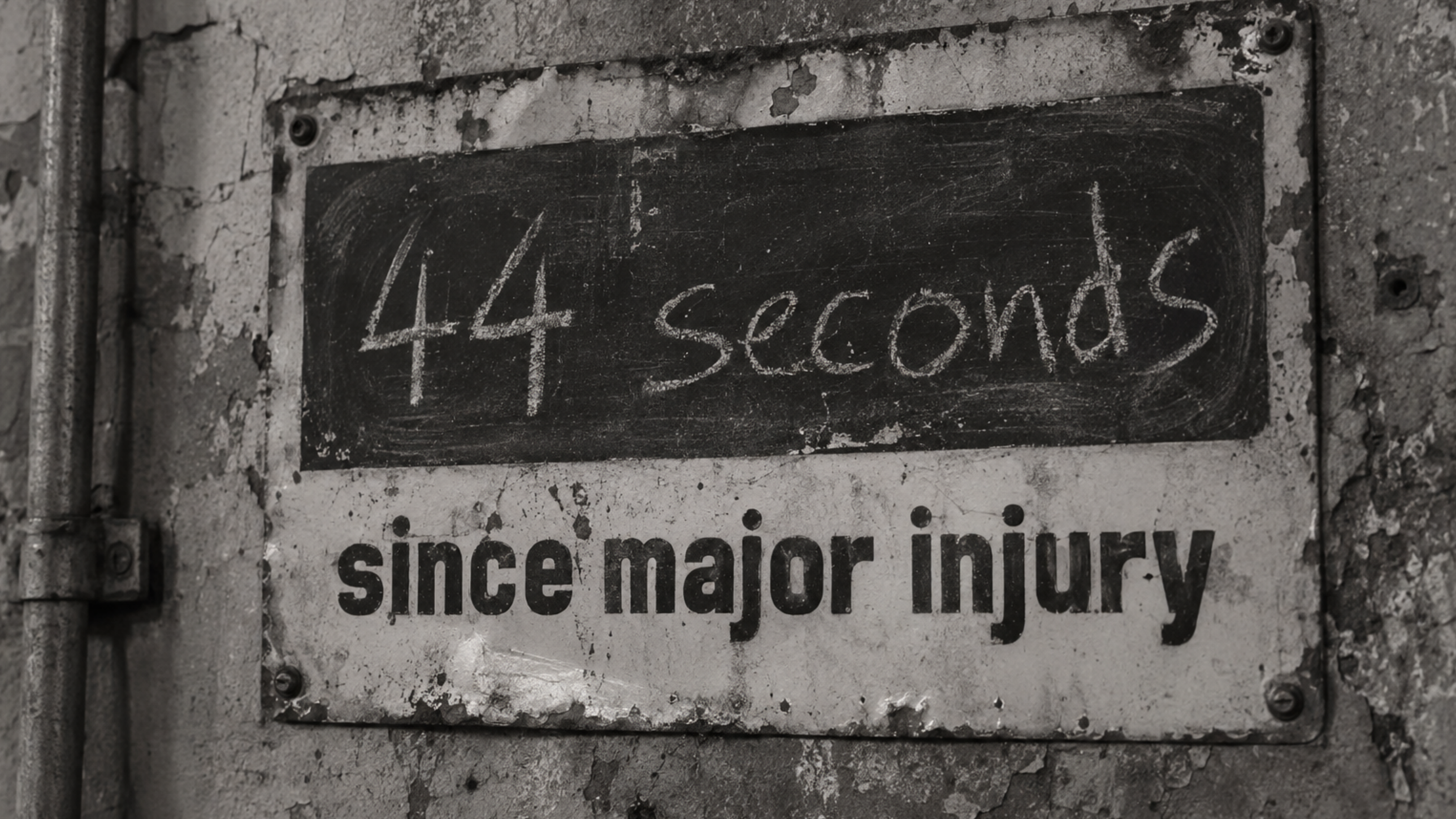
Big, heavy, expensive, steam engine.

@jpschroed








44 seconds
since major injury

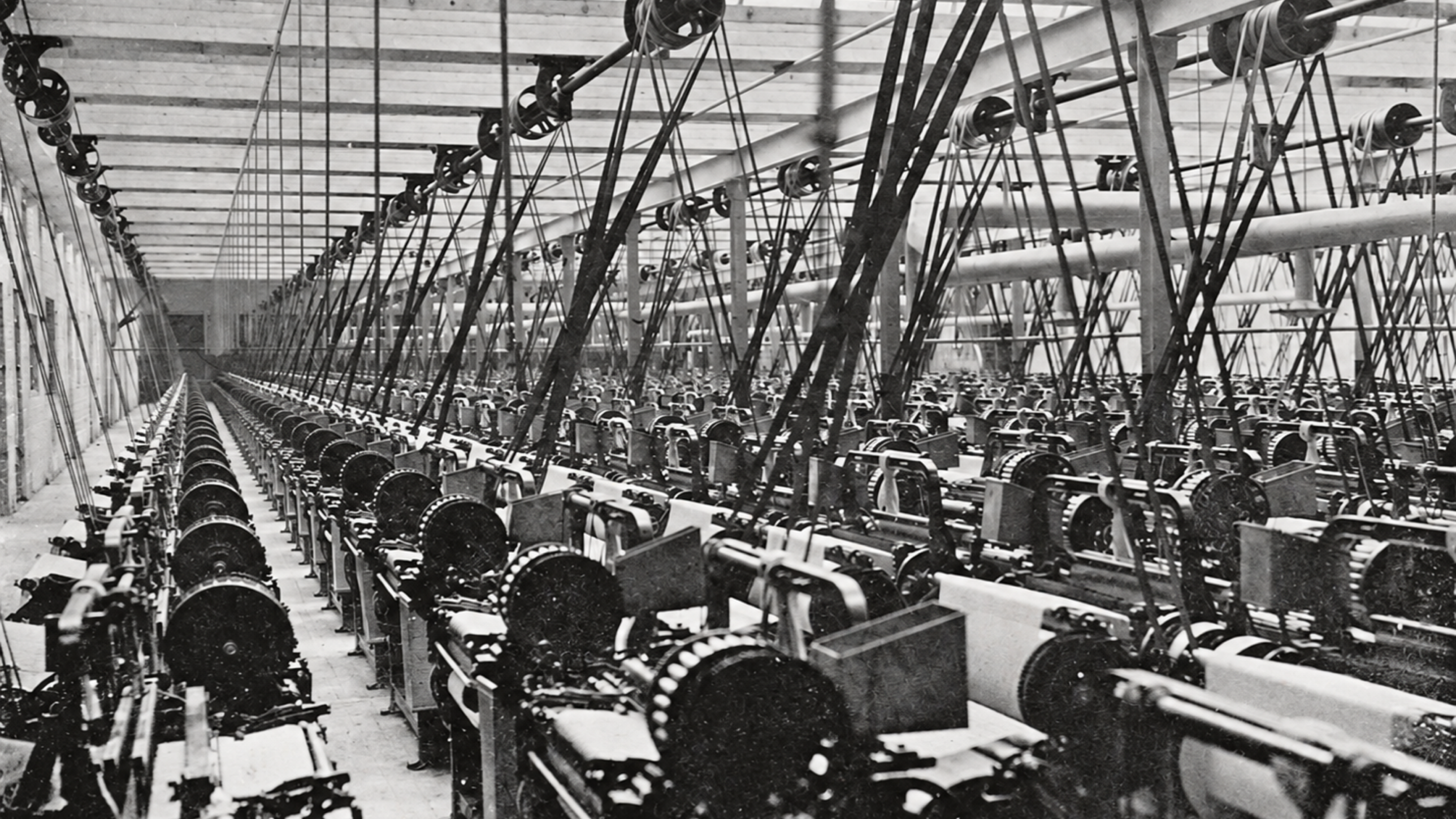

You Are Here

We are in the line shaft era of agents.

Actual things people need to do



Big, heavy, expensive, model.





How?

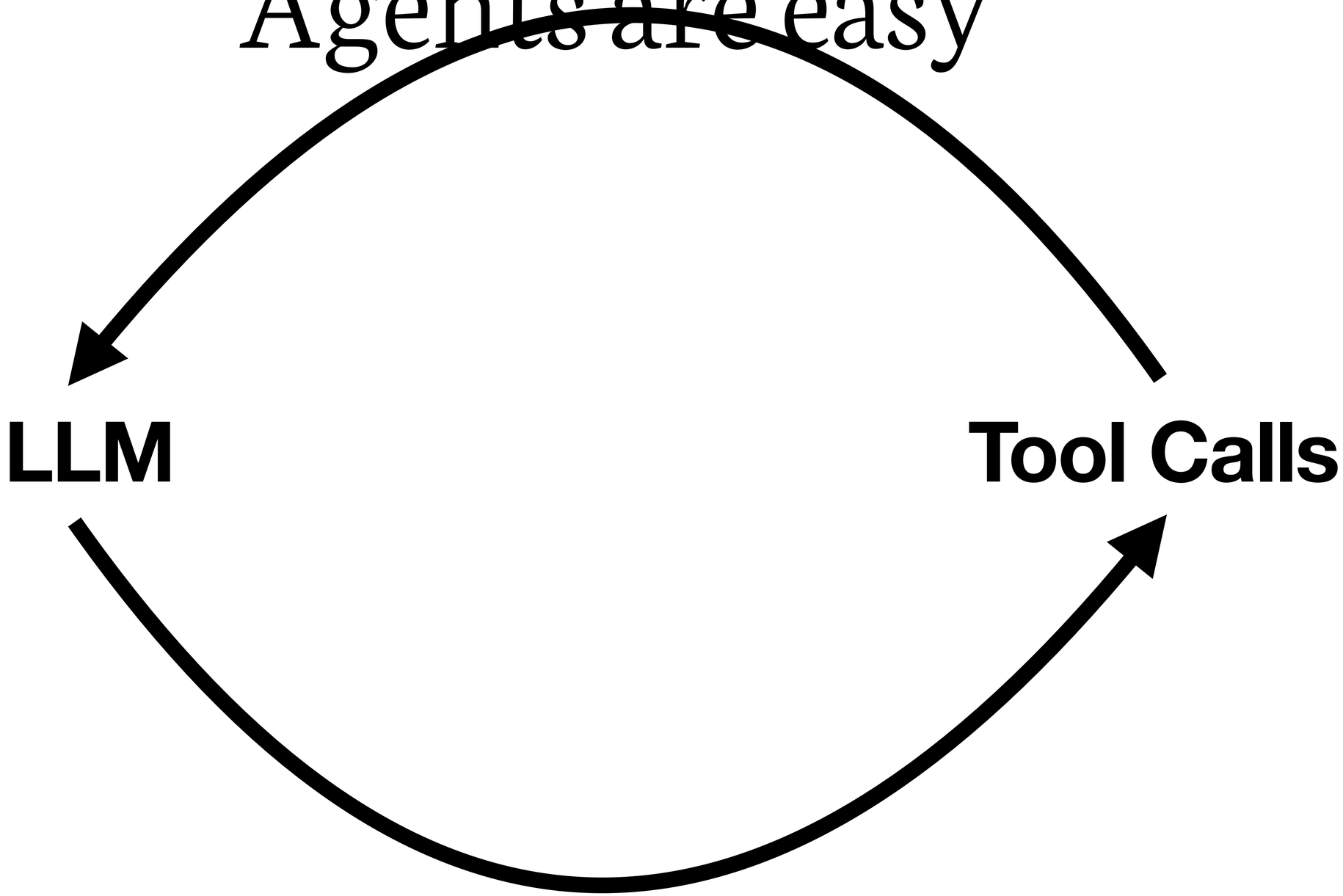
What is an Agent?

Agent = Harness

A harness is *deterministic* software that utilizes the non-deterministic results produced by models in pursuit of a desired objective.

Agent = Harness = Software

~~Agents are~~ easy



@jpschroeder

Agents are hard

- Agentic loop orchestration
- Provider abstraction
- Durable execution
- Tool call validation
- Stop conditions
- Multi-agent turn coordination
- Durable thread state
- Message hierarchy (sub-prompts)
- Context window management
- Parallel tool execution
- Real-time log streaming
- Memory-efficient image processing
- Large file chunking
- Automatic image optimization
- Zero-config discovery
- Message lifecycle hooks
- Tool lifecycle hooks
- Framework-agnostic client
- Workblock groupings
- Agent packaging & distribution
- Sub-prompt chaining
- Graceful execution abort
- Serial tool execution
- Retry with backoff
- Unified response format
- HTTP streaming for long execution
- WebSocket persistence
- Execution telemetry
- Namespace isolation
- Hallucinated tool rejection

~~Agents~~



But *everyone* is building agents

1. Integration
2. Trust
3. Efficacy

Integration

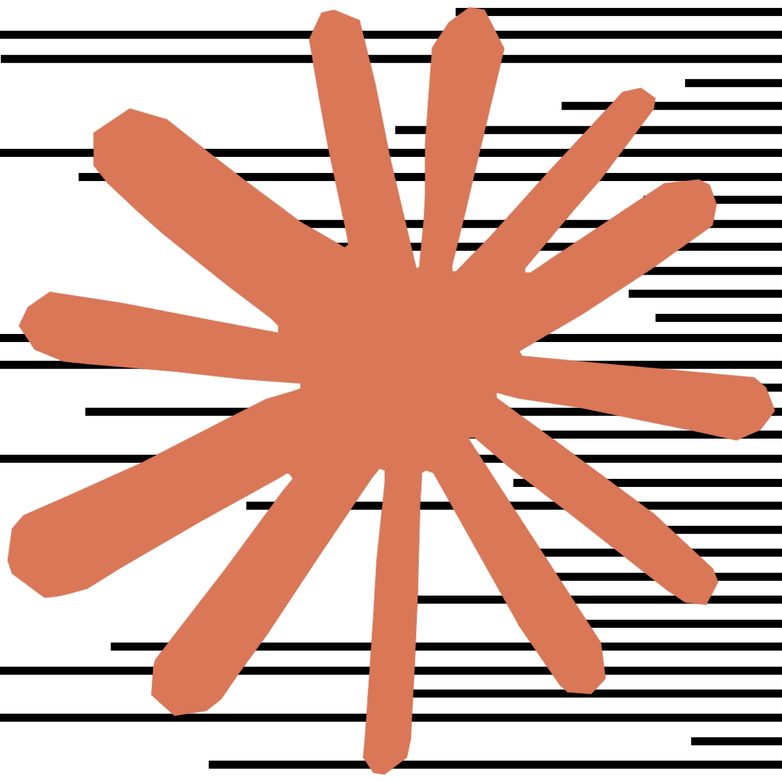
> If I hand wrote my code on a napkin, then tore it into little pieces, put each one in a different envelope, and mailed them all over the world, then back to myself, then manually put the pieces back together, and typed it into the computer….

That still wouldn’t be the bottleneck to shipping new features.

Integration

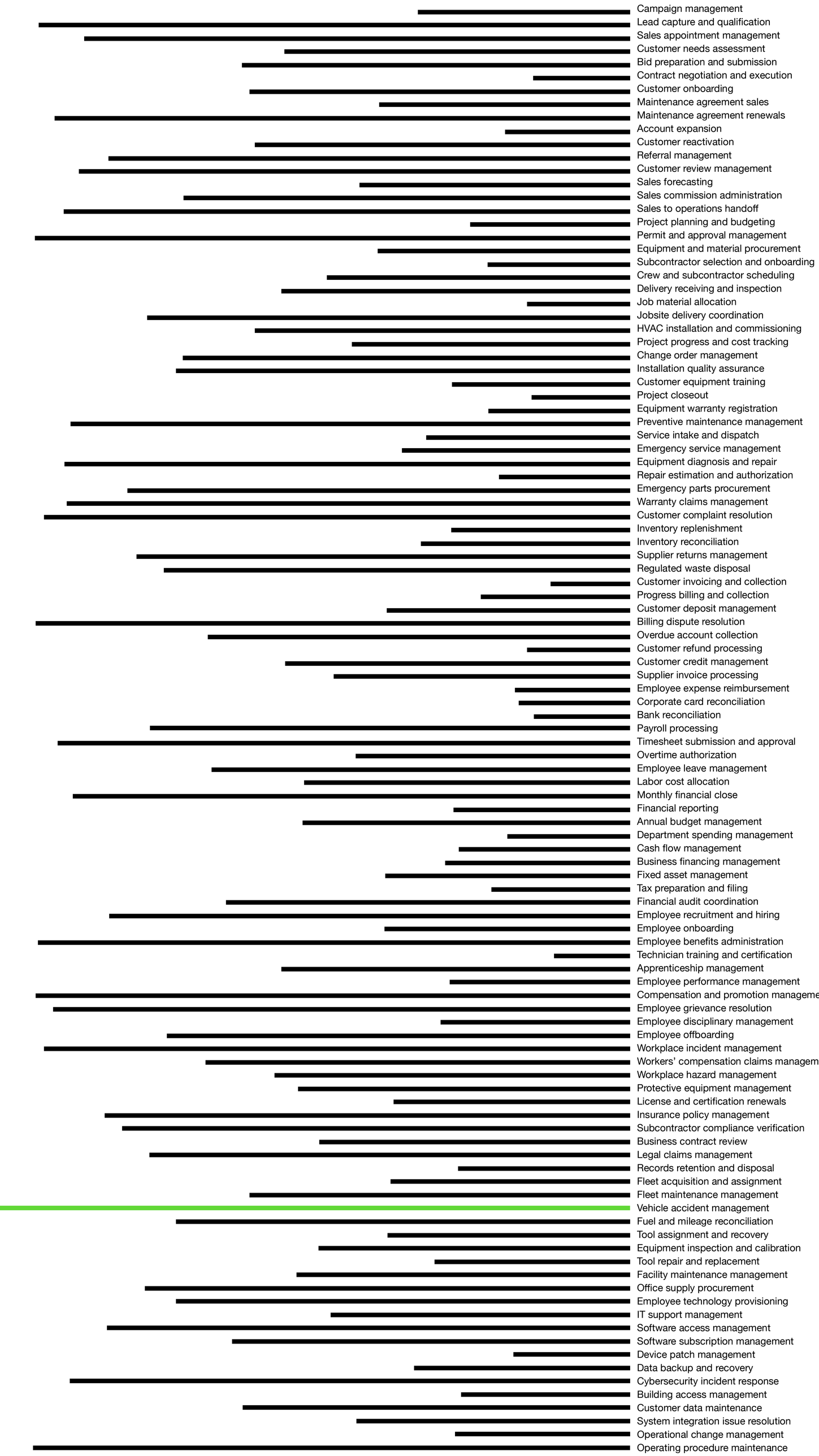


- Campaign management
- Lead capture and qualification
- Sales appointment management
- Customer needs assessment
- Bid preparation and submission
- Contract negotiation and execution
- Customer onboarding

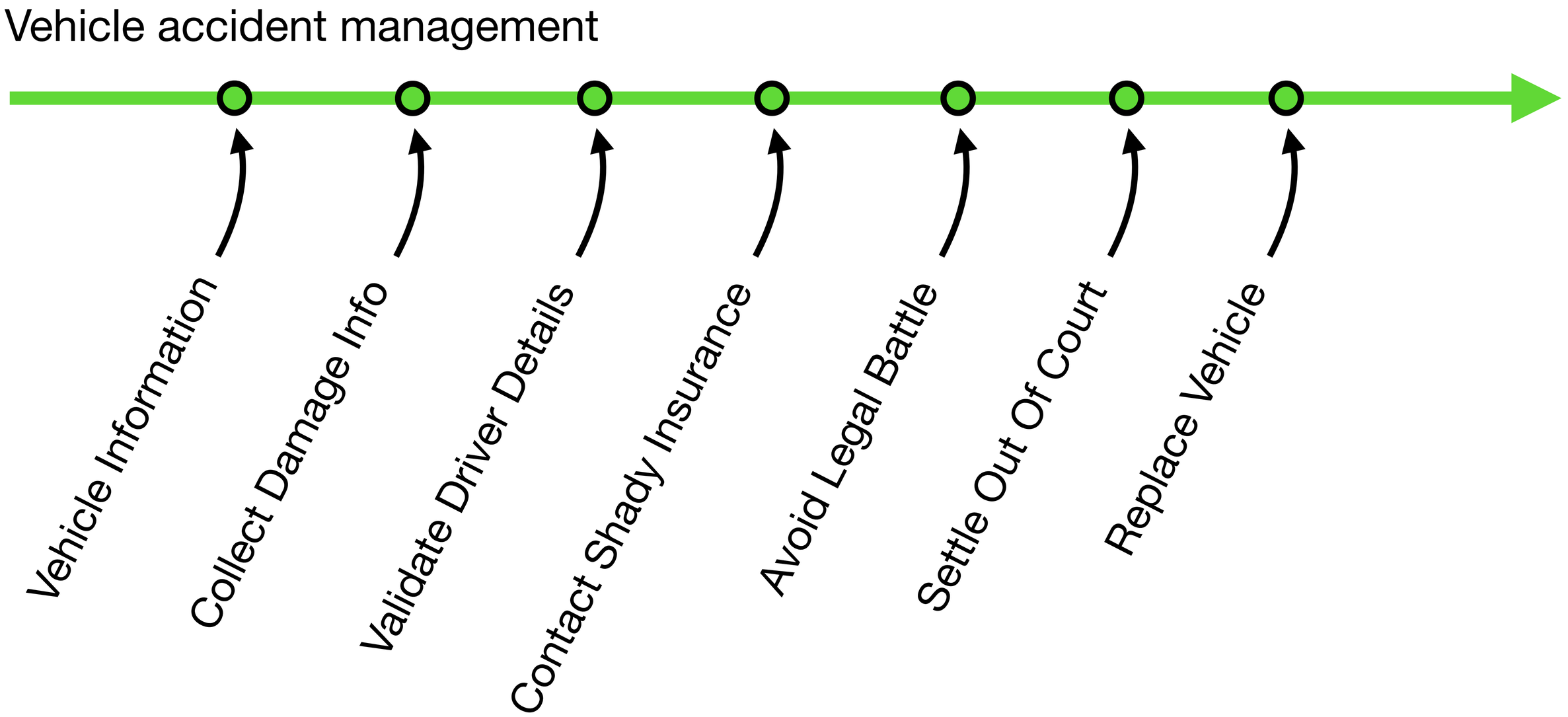


Campaign management
Lead capture and qualification
Sales appointment management
Customer needs assessment
Bid preparation and submission
Contract negotiation and execution
Customer onboarding
Maintenance agreement sales
Maintenance agreement renewals
Account expansion
Customer reactivation
Referral management
Customer review management
Sales forecasting
Sales commission administration
Sales to operations handoff
Project planning and budgeting
Permit and approval management
Equipment and material procurement
Subcontractor selection and onboarding
Crew and subcontractor scheduling
Delivery receiving and inspection
Job material allocation
Jobsite delivery coordination
HVAC installation and commissioning
Project progress and cost tracking
Change order management
Installation quality assurance
Customer equipment training
Project closeout
Equipment warranty registration
Preventive maintenance management
Service intake and dispatch
Emergency service management
Equipment diagnosis and repair
Repair estimation and authorization
Emergency parts procurement
Warranty claims management
Customer complaint resolution
Inventory replenishment
Inventory reconciliation
Supplier returns management
Regulated waste disposal
Customer invoicing and collection
Progress billing and collection
Customer deposit management
Billing dispute resolution
Overdue account collection
Customer refund processing
Customer credit management
Supplier invoice processing
Employee expense reimbursement
Corporate card reconciliation
Bank reconciliation
Payroll processing
Timesheet submission and approval
Overtime authorization
Employee leave management
Labor cost allocation
Monthly financial close
Financial reporting
Annual budget management
Department spending management
Cash flow management
Business financing management
Fixed asset management
Tax preparation and filing
Financial audit coordination
Employee recruitment and hiring
Employee onboarding
Employee benefits administration
Technician training and certification
Apprenticeship management
Employee performance management
Compensation and promotion manageme
Employee grievance resolution
Employee disciplinary management
Employee offboarding
Workplace incident management
Workers' compensation claims managem
Workplace hazard management
Protective equipment management
License and certification renewals
Insurance policy management
Subcontractor compliance verification
Business contract review
Legal claims management
Records retention and disposal
Fleet acquisition and assignment
Fleet maintenance management
Vehicle accident management
Fuel and mileage reconciliation
Tool assignment and recovery
Equipment inspection and calibration
Tool repair and replacement
Facility maintenance management
Office supply procurement
Employee technology provisioning
IT support management
Software access management
Software subscription management
Device patch management
Data backup and recovery
Cybersecurity incident response
Building access management
Customer data maintenance
System integration issue resolution
Operational change management
Operating procedure maintenance

It doesn’t even matter if it works.

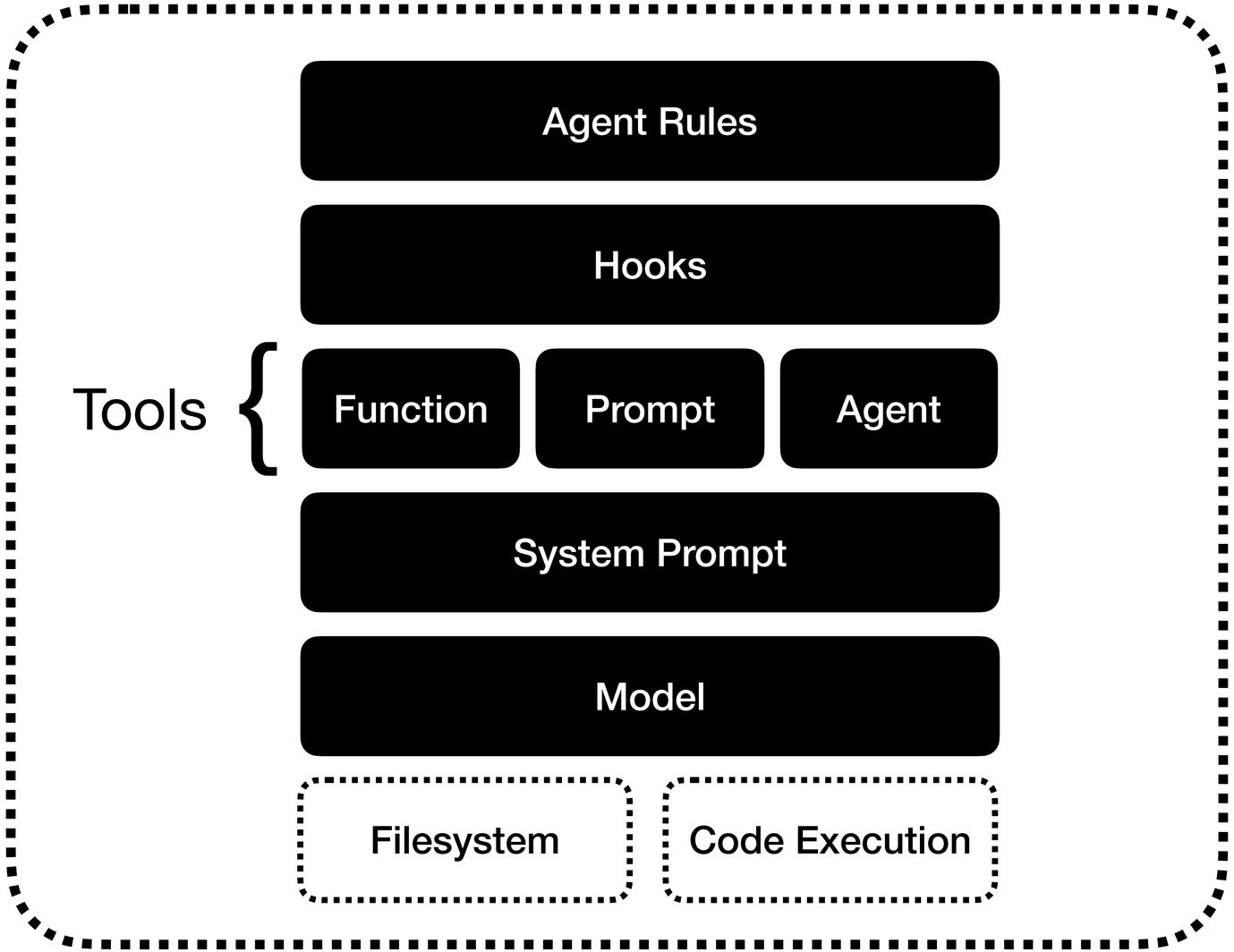

Campaign management
Lead capture and qualification
Sales appointment management
Customer needs assessment
Bid preparation and submission
Contract negotiation and execution
Customer onboarding
Maintenance agreement sales
Maintenance agreement renewals
Account expansion
Customer reactivation
Referral management
Customer review management
Sales forecasting
Sales commission administration
Sales to operations handoff
Project planning and budgeting
Permit and approval management
Equipment and material procurement
Subcontractor selection and onboarding
Crew and subcontractor scheduling
Delivery receiving and inspection
Job material allocation
Jobsite delivery coordination
HVAC installation and commissioning
Project progress and cost tracking
Change order management
Installation quality assurance
Customer equipment training
Project closeout
Equipment warranty registration
Preventive maintenance management
Service intake and dispatch
Emergency service management
Equipment diagnosis and repair
Repair estimation and authorization
Emergency parts procurement
Warranty claims management
Customer complaint resolution
Inventory replenishment
Inventory reconciliation
Supplier returns management
Regulated waste disposal
Customer invoicing and collection
Progress billing and collection
Customer deposit management
Billing dispute resolution
Overdue account collection
Customer refund processing
Customer credit management
Supplier invoice processing
Employee expense reimbursement
Corporate card reconciliation
Bank reconciliation
Payroll processing
Timesheet submission and approval
Overtime authorization
Employee leave management
Labor cost allocation
Monthly financial close
Financial reporting
Annual budget management
Department spending management
Cash flow management
Business financing management
Fixed asset management
Tax preparation and filing
Financial audit coordination
Employee recruitment and hiring
Employee onboarding
Employee benefits administration
Technician training and certification
Apprenticeship management
Employee performance management
Compensation and promotion manageme
Employee grievance resolution
Employee disciplinary management
Employee offboarding
Workplace incident management
Workers’ compensation claims managem
Workplace hazard management
Protective equipment management
License and certification renewals
Insurance policy management
Subcontractor compliance verification
Business contract review
Legal claims management
Records retention and disposal
Fleet acquisition and assignment
Fleet maintenance management
Vehicle accident management
Fuel and mileage reconciliation
Tool assignment and recovery
Equipment inspection and calibration
Tool repair and replacement
Facility maintenance management
Office supply procurement
Employee technology provisioning
IT support management
Software access management
Software subscription management
Device patch management
Data backup and recovery
Cybersecurity incident response
Building access management
Customer data maintenance
System integration issue resolution
Operational change management
Operating procedure maintenance

Integration

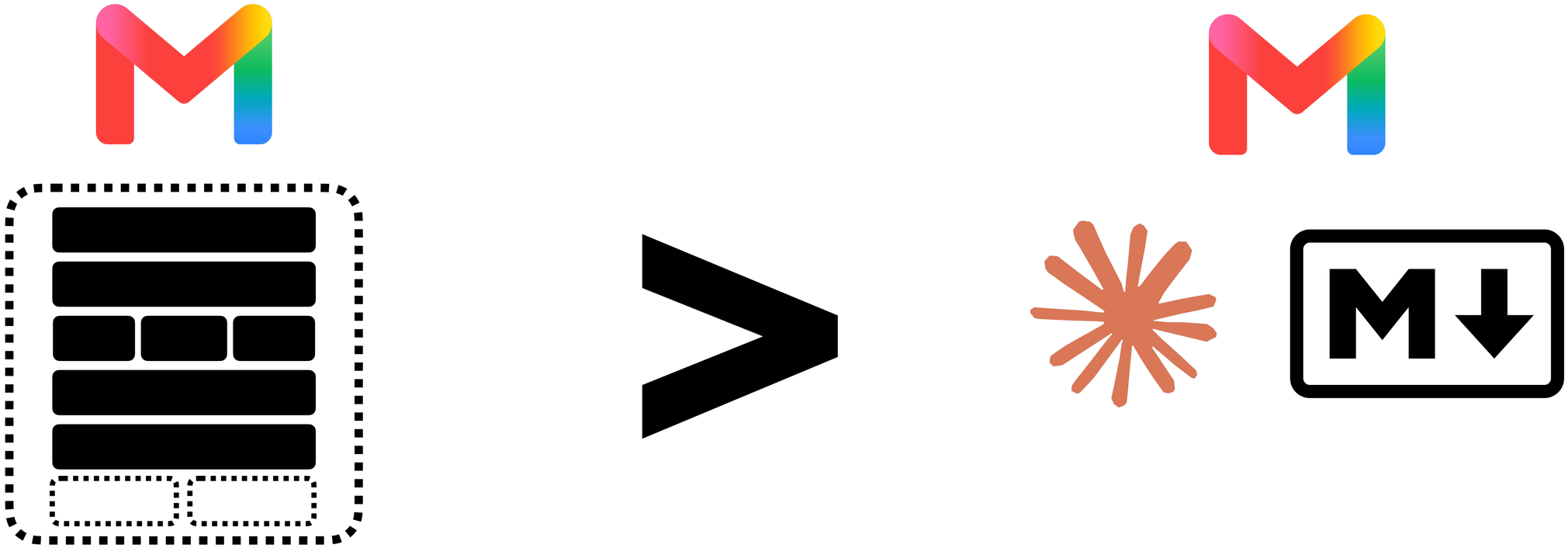


1. Integration
2. Trust
3. Efficacy

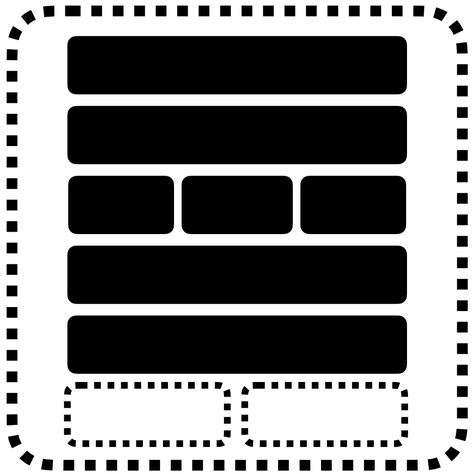
Trust


Agent Rules
Hooks
Tools
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code Execution

Trust



Trust

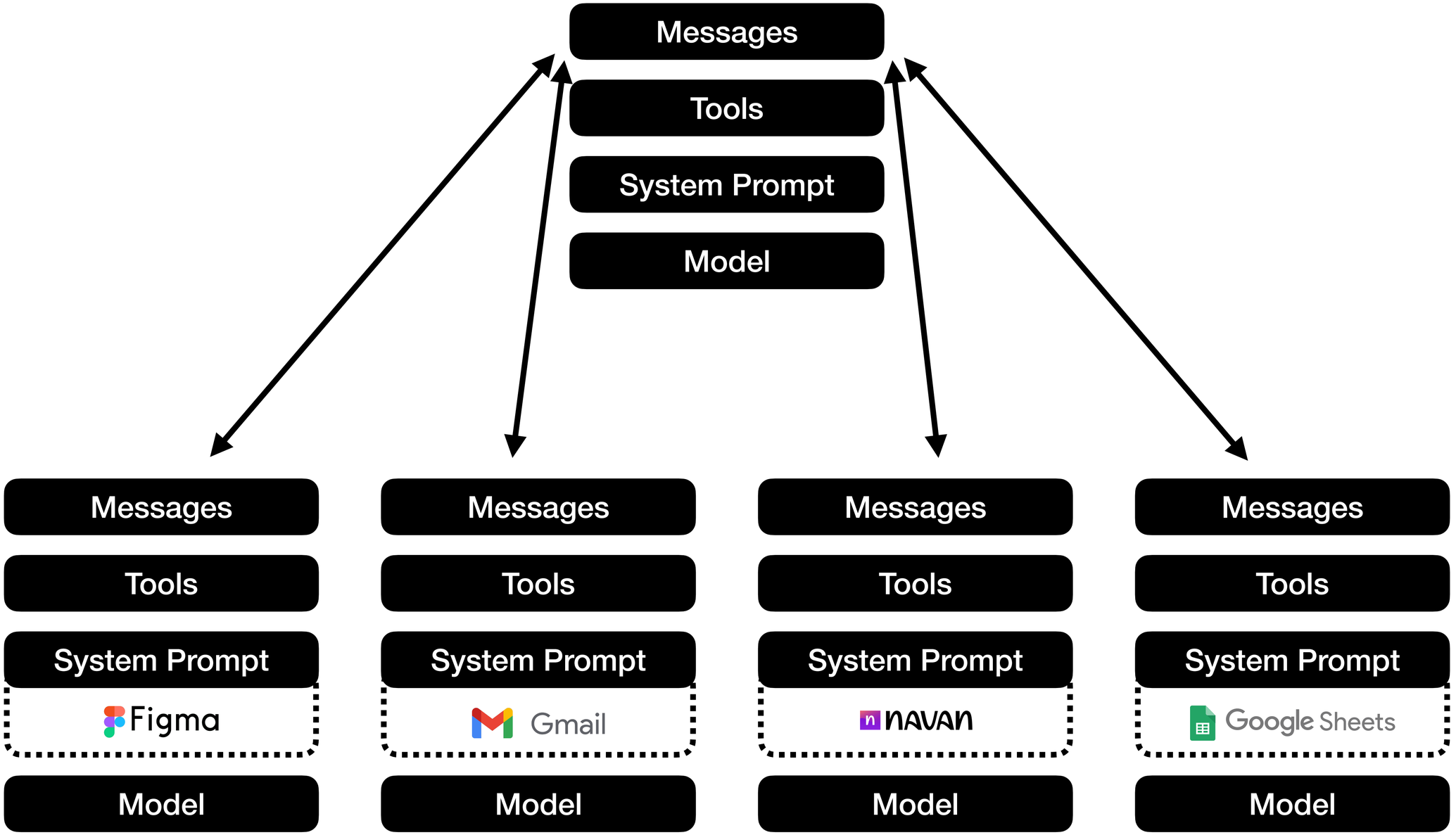


Go check twitter please

Sorry, I can’t know how to do that

Go check twitter please

All software has been hacked. Shutting down prod now as a preventative measure.


Messages
Tools
System Prompt
Model
Messages
Tools
System Prompt
Figma
Model
Messages
Tools
System Prompt
Gmail
Model
Messages
Tools
System Prompt
NAVAN
Model
Messages
Tools
System Prompt
Google Sheets
Model

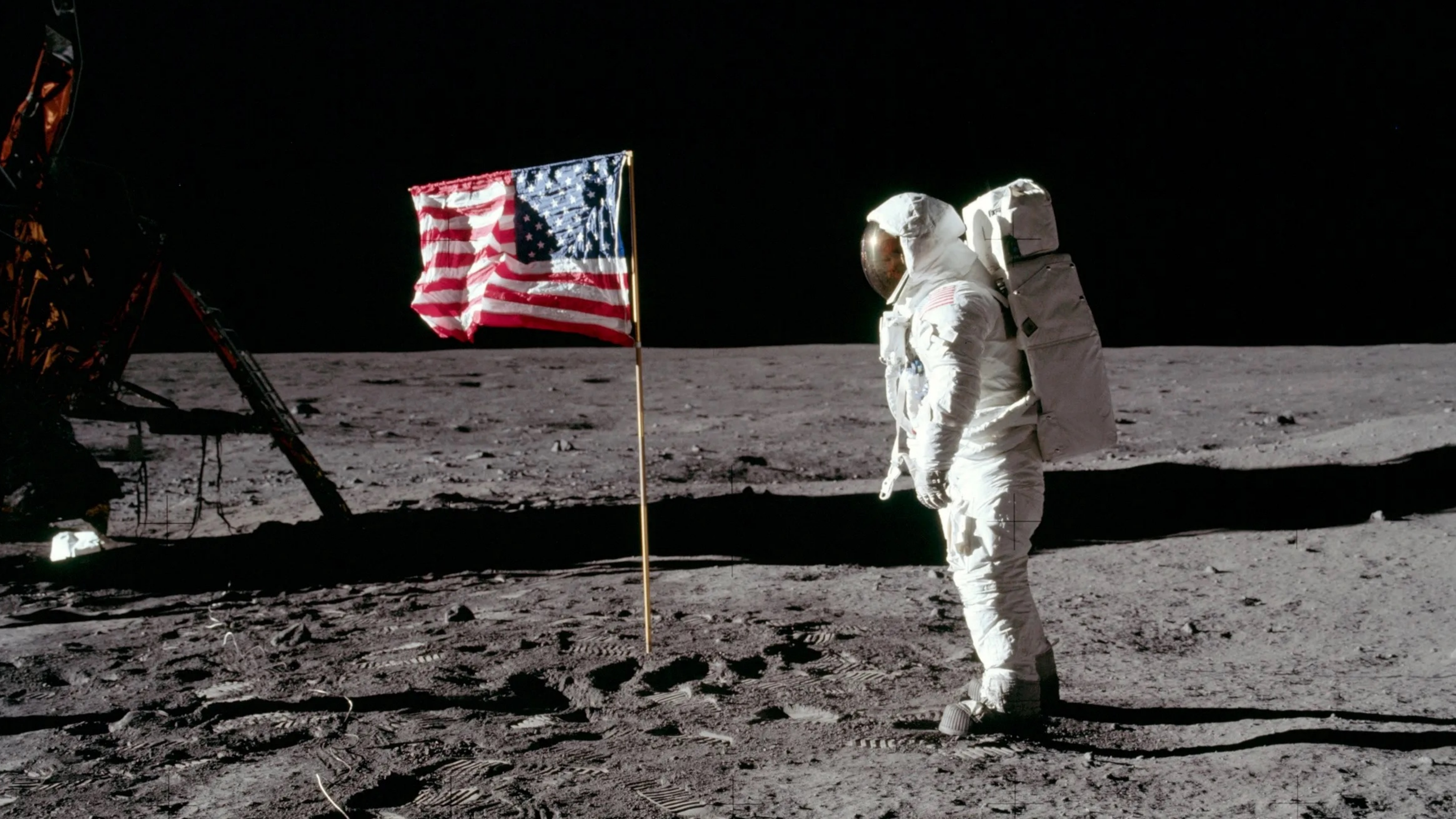
It works...

Principle of least privilege

Law of Demeter

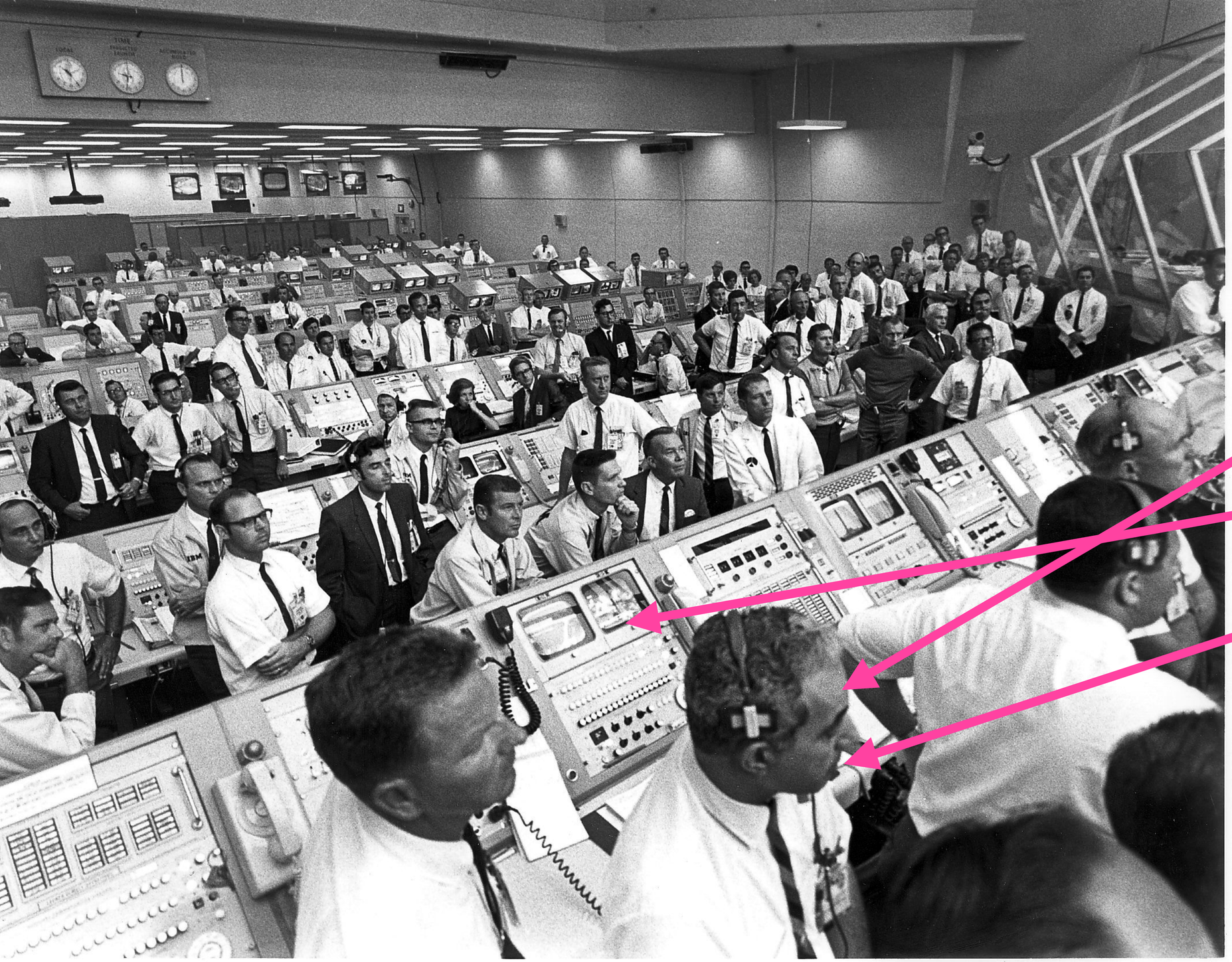
Composition over inheritance

Separation of concerns





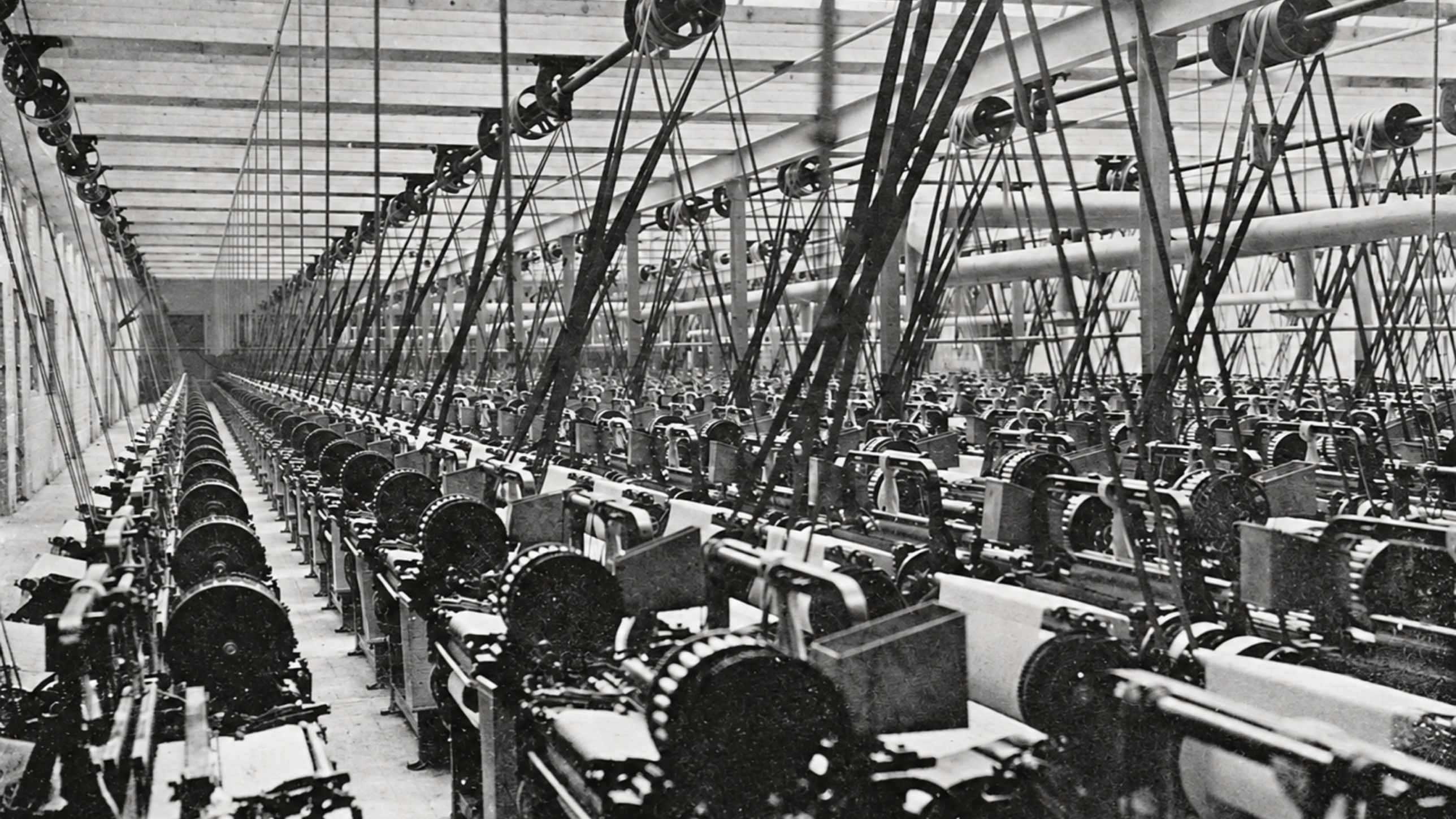


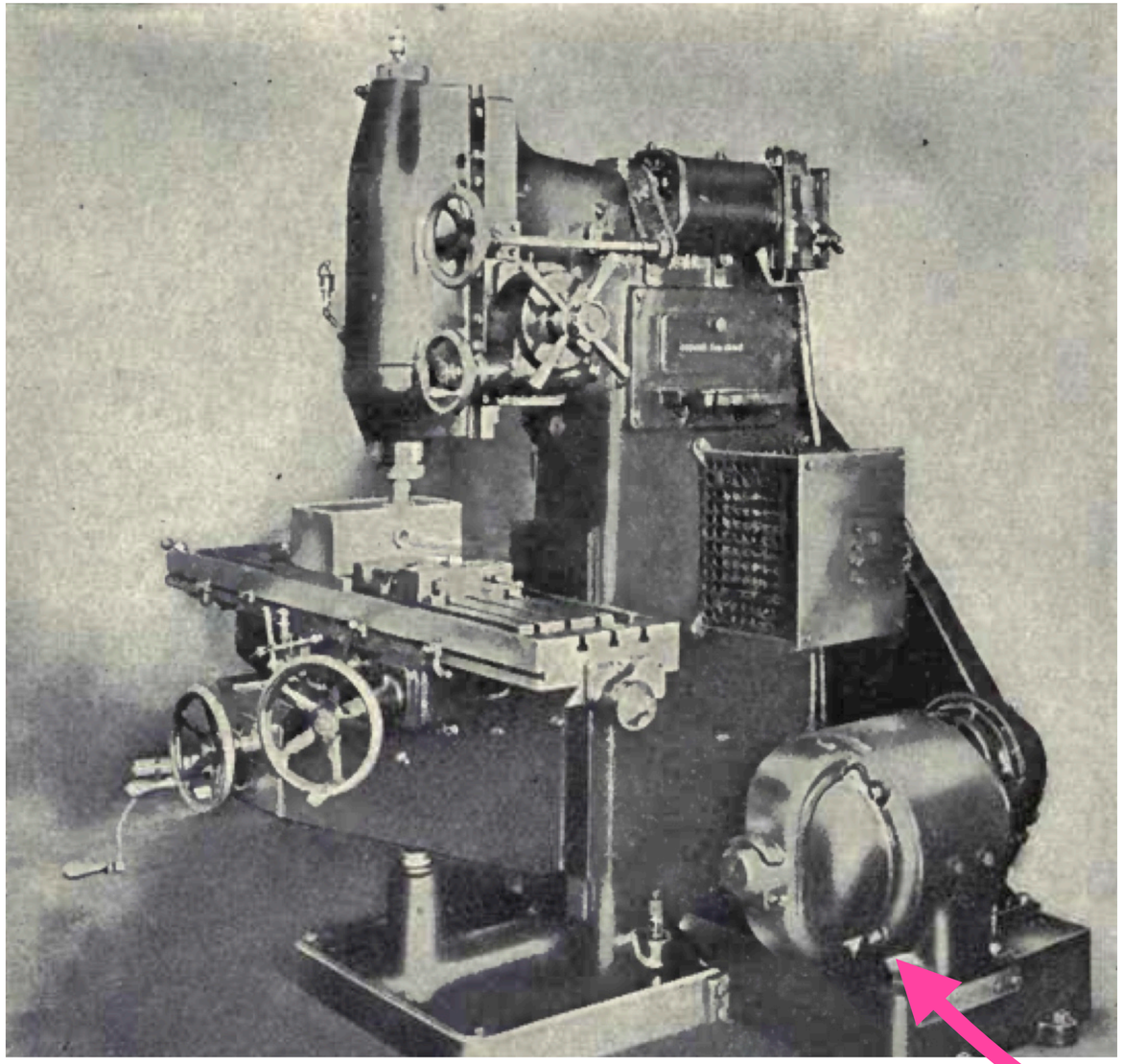


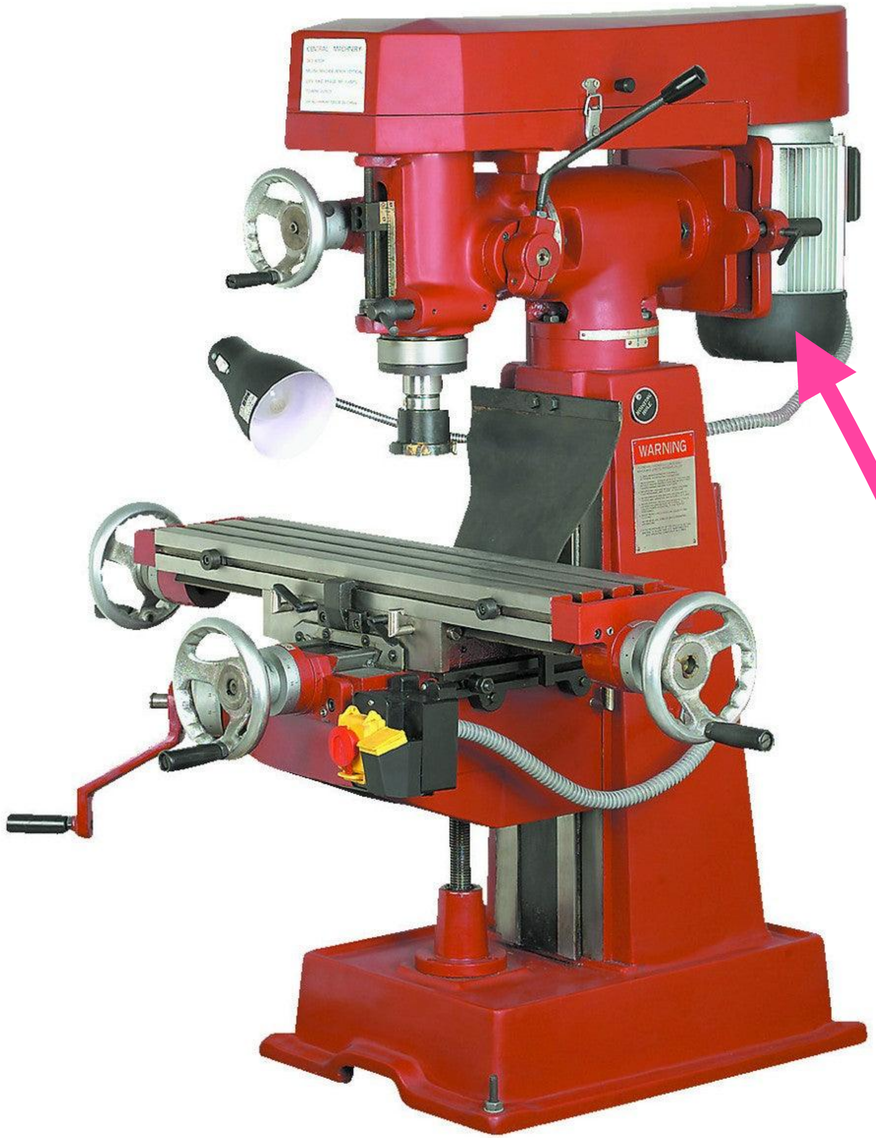
Agent

Tools

Messages

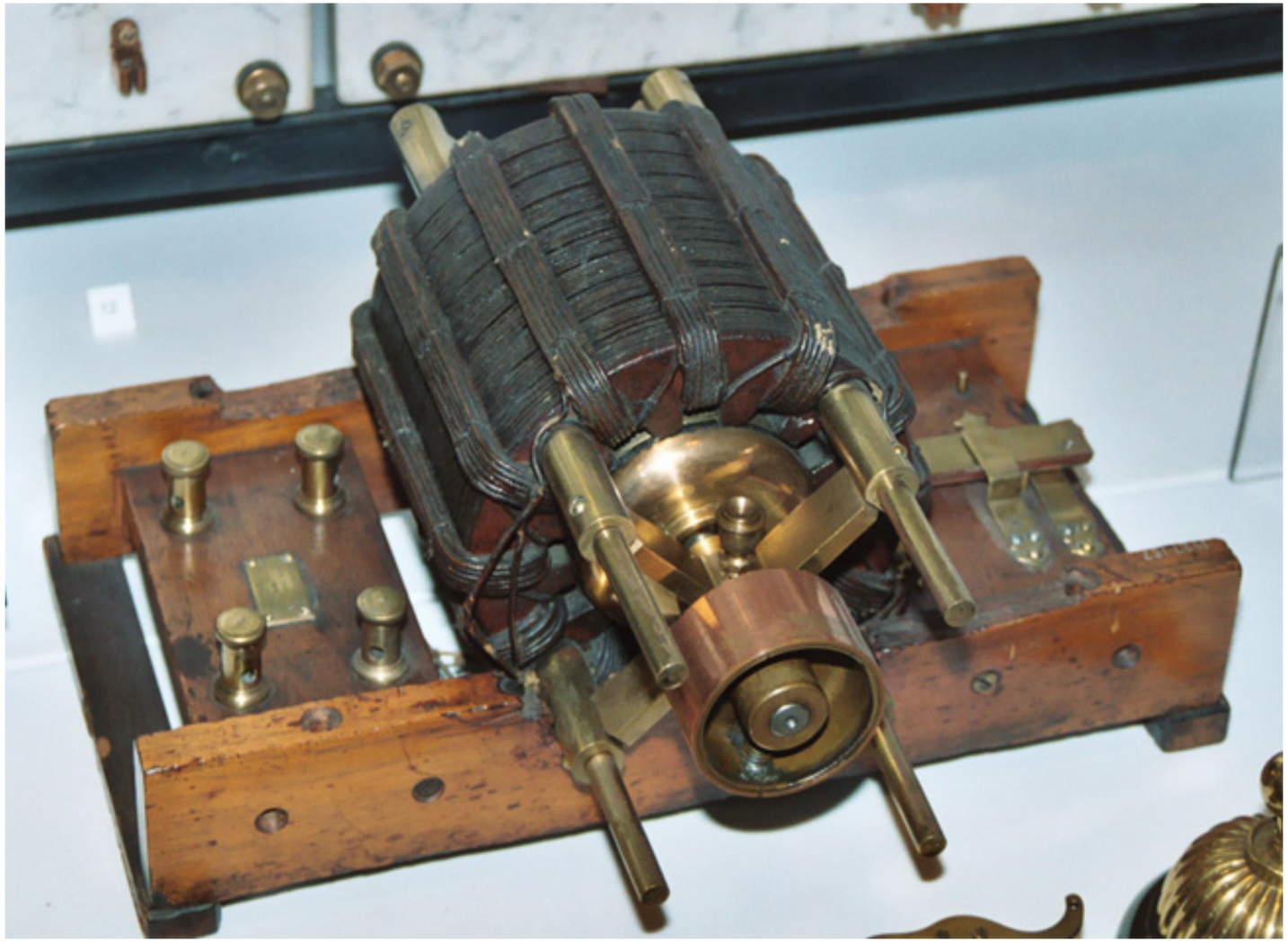






@jpschroeder

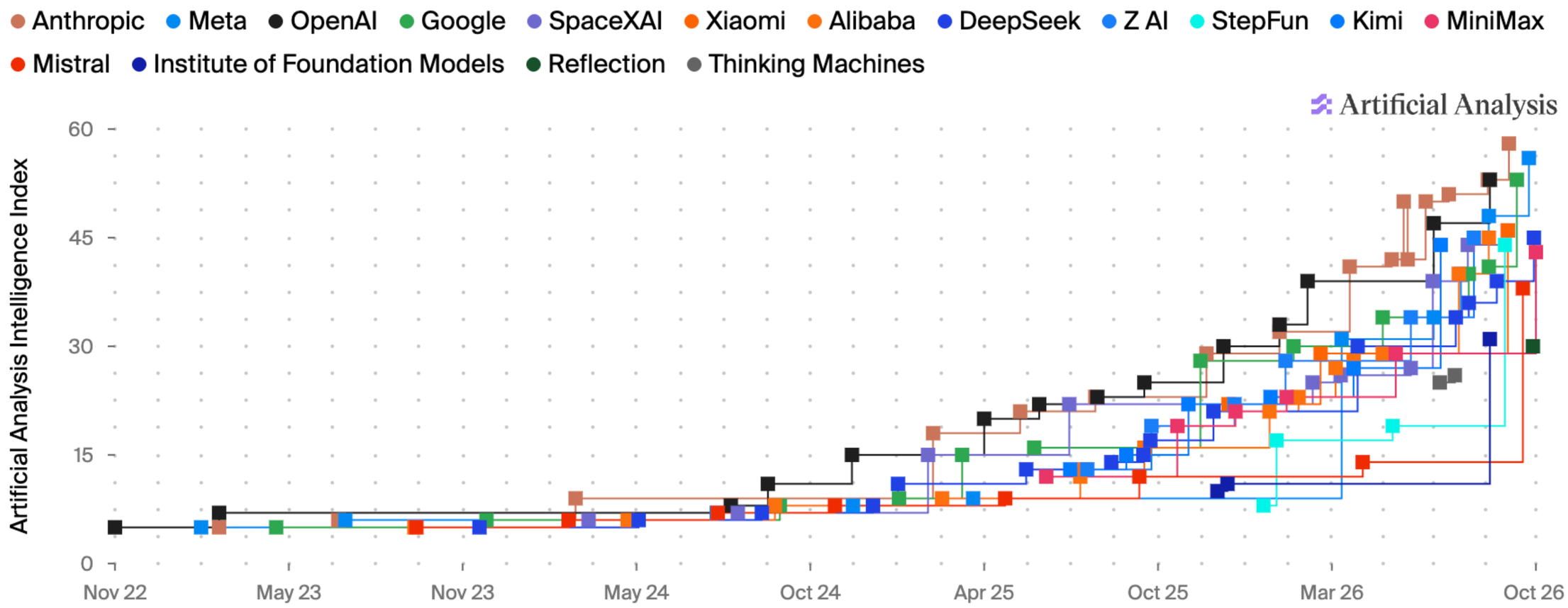
Electric motors!



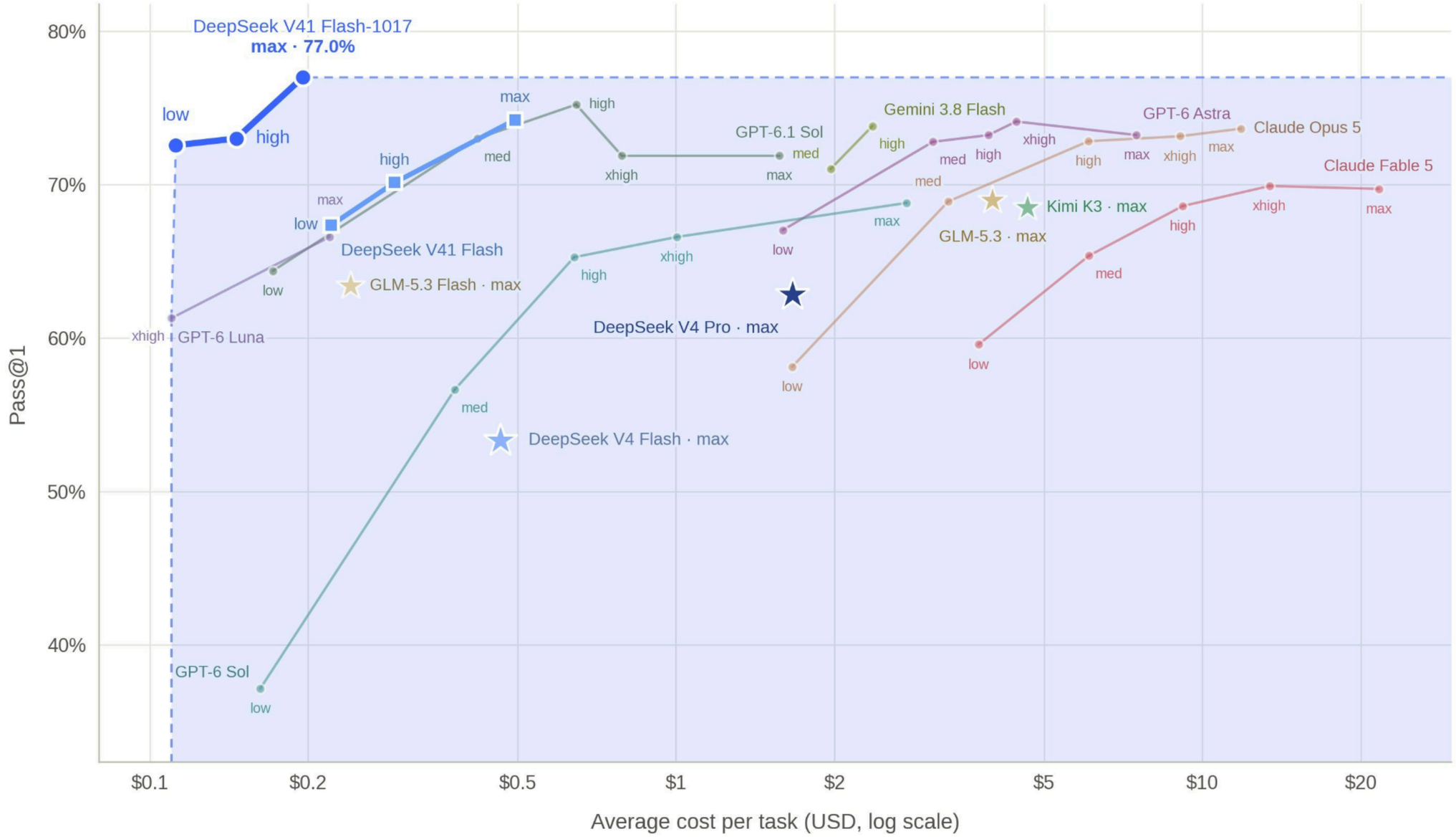
@jpschroeder

1. Integration
2. Trust
3. Efficacy

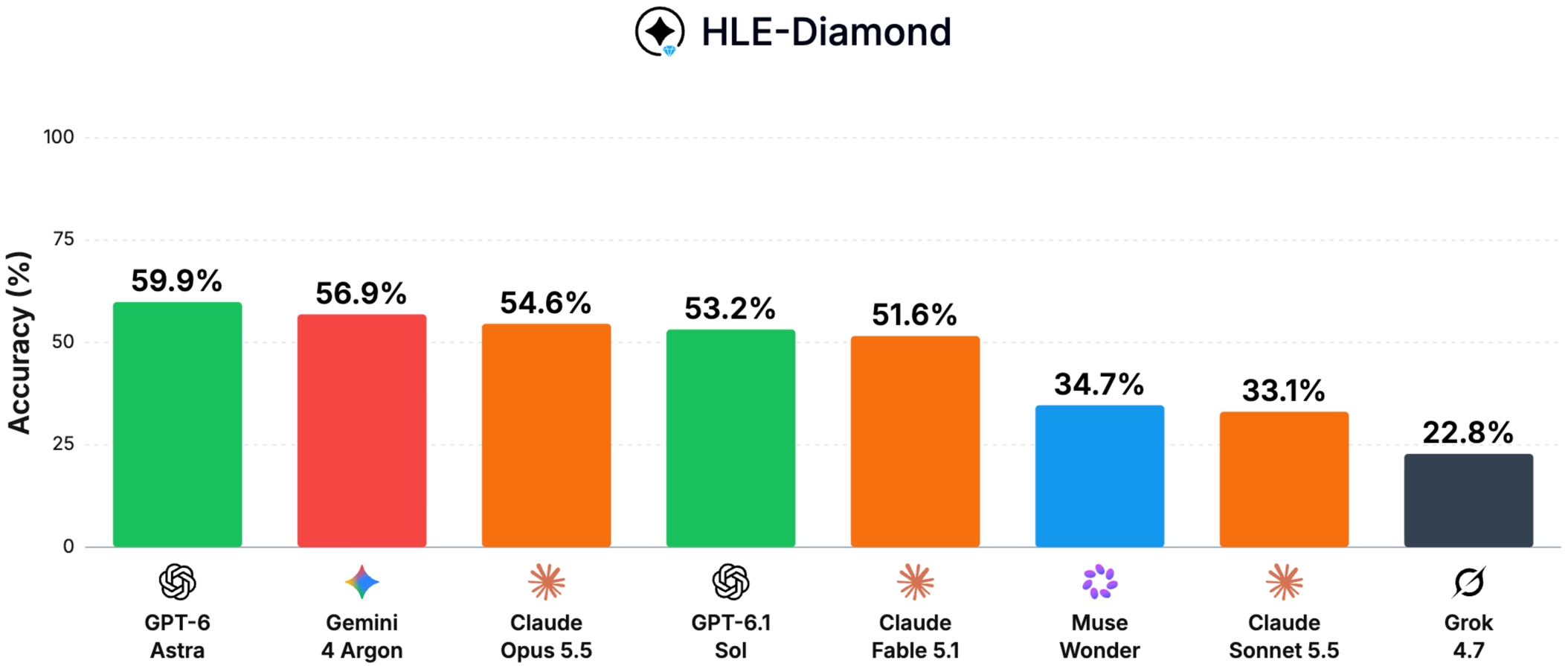
Efficacy


Anthropic
Meta
OpenAI
Google
SpaceXAI
Xiaomi
Alibaba
DeepSeek
Z AI
StepFun
Kimi
MiniMax
Mistral
Institute of Foundation Models
Reflection
Thinking Machines
Artificial Analysis
Artificial Analysis Intelligence Index
60
45
30
15
0
Nov 22
May 23
Nov 23
May 24
Oct 24
Apr 25
Oct 25
Mar 26
Oct 26

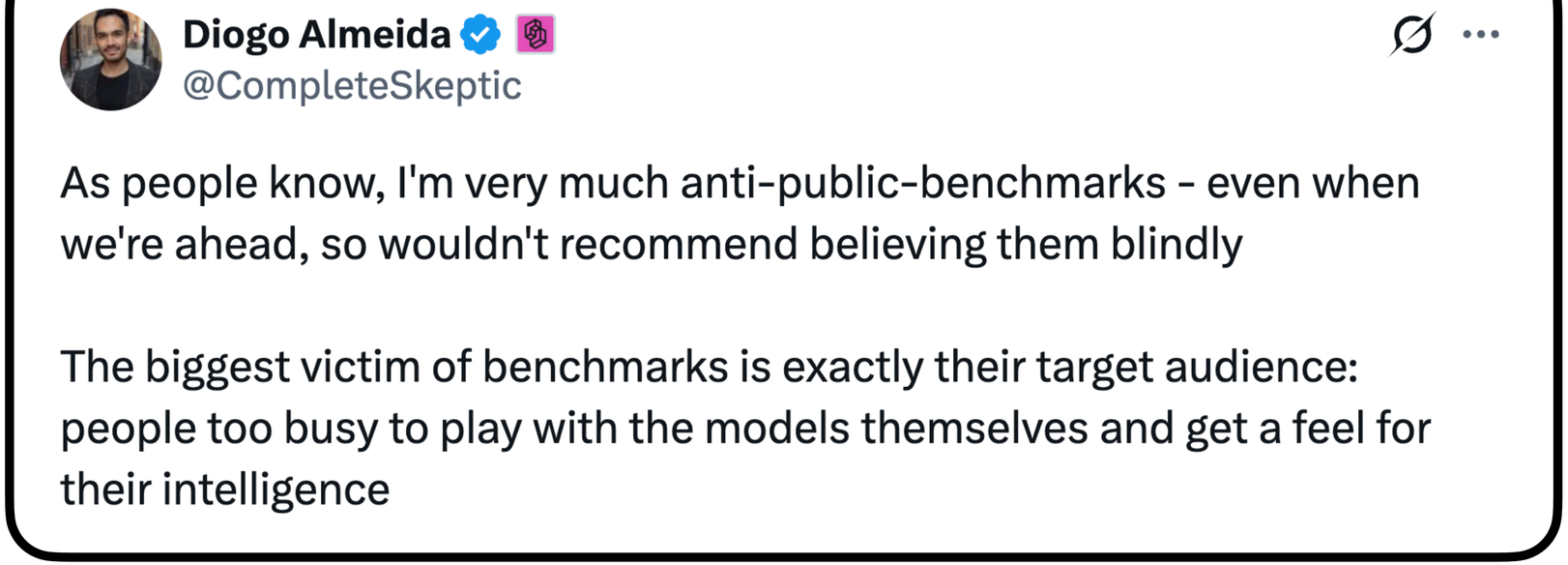
Efficacy


DeepSeek V41 Flash-1017
max · 77.0%
low
high
max
high
med
xhigh
max
low
DeepSeek V41 Flash
GLM-5.3 Flash · max
low
xhigh
GPT-6 Luna
high
xhigh
max
GPT-6.1 Sol
med
max
Gemini 3.8 Flash
high
med
high
xhigh
GPT-6 Astra
high
max
xhigh
max
Claude Opus 5
med
low
DeepSeek V4 Pro · max
GLM-5.3 · max
Kimi K3 · max
low
med
high
xhigh
max
Claude Fable 5
med
DeepSeek V4 Flash · max
GPT-6 Sol
low
80%
70%
60%
50%
40%
Pass@1
$0.1
$0.2
$0.5
$1
$2
$5
$10
$20
Average cost per task (USD, log scale)

Efficacy


HLE-Diamond
Accuracy (%)
100
75
50
25
0
59.9%
56.9%
54.6%
53.2%
51.6%
34.7%
33.1%
22.8%
GPT-6 Astra
Gemini 4 Argon
Claude Opus 5.5
GPT-6.1 Sol
Claude Fable 5.1
Muse Wonder
Claude Sonnet 5.5
Grok 4.7

Efficacy

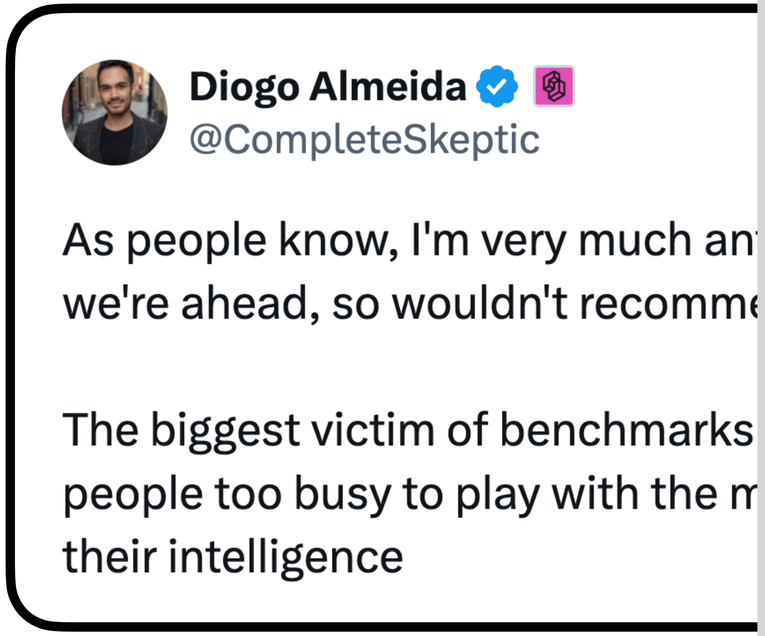


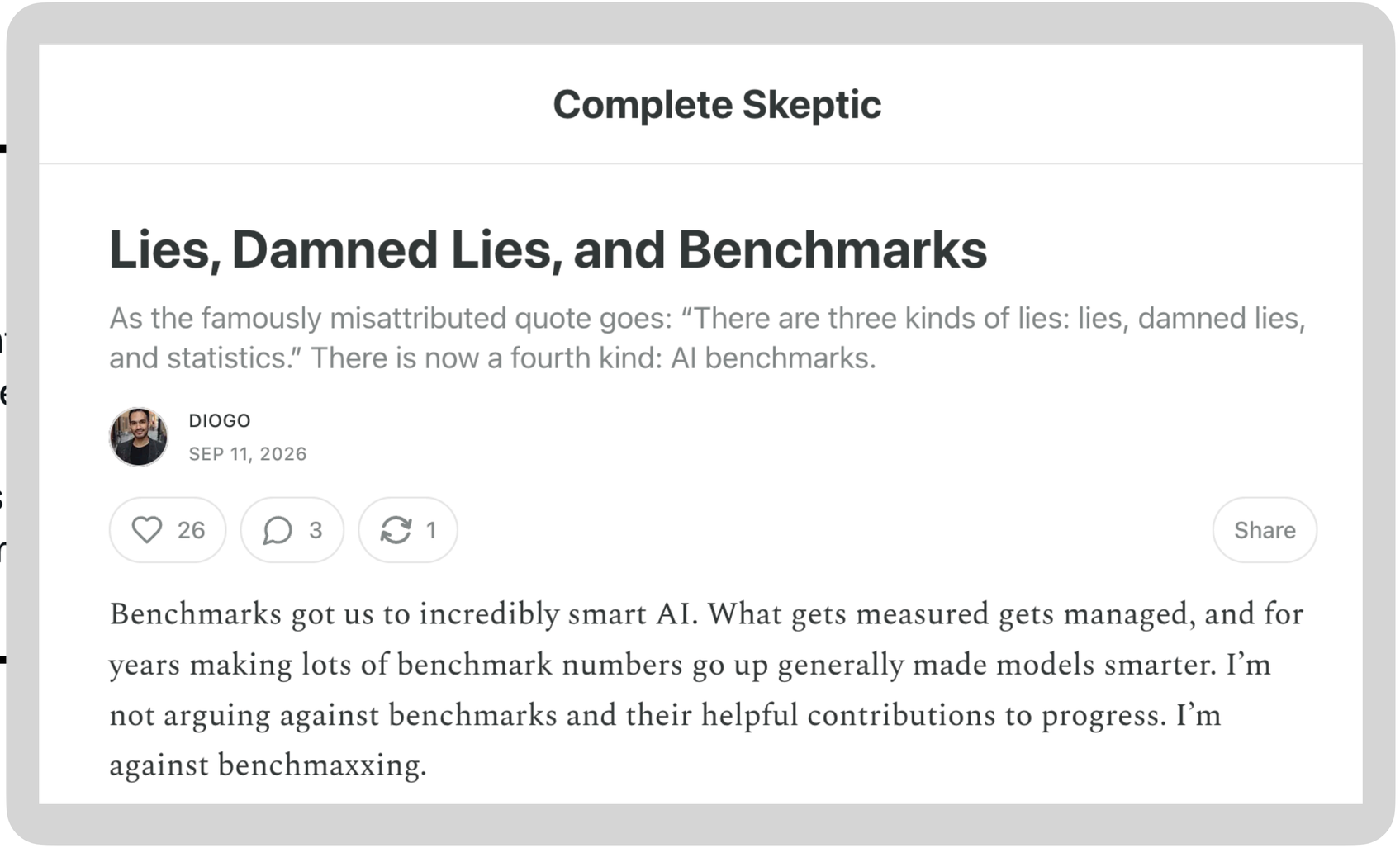
Diogo Almeida
@CompleteSkeptic

As people know, I'm very much anti-public-benchmarks - even when we're ahead, so wouldn't recommend believing them blindly

The biggest victim of benchmarks is exactly their target audience: people too busy to play with the models themselves and get a feel for their intelligence

Efficacy


Diogo Almeida
@CompleteSkeptic
As people know, I'm very much an
we're ahead, so wouldn't recomme
The biggest victim of benchmarks
people too busy to play with the m
their intelligence


Complete Skeptic
Lies, Damned Lies, and Benchmarks
As the famously misattributed quote goes: "There are three kinds of lies: lies, damned lies, and statistics." There is now a fourth kind: AI benchmarks.
DIOGO
SEP 11, 2026
26
3
1
Share
Benchmarks got us to incredibly smart AI. What gets measured gets managed, and for years making lots of benchmark numbers go up generally made models smarter. I'm not arguing against benchmarks and their helpful contributions to progress. I'm against benchmaxxing.



Streetlight Effect

Efficacy

“The models are jagged”



A LONG-HORIZON LLM BENCHMARK

Model Zen Garden

explore two gardens · vote · view results

Efficacy

- Far more efficient with tokens

“Get the last email from Debby”



Efficacy

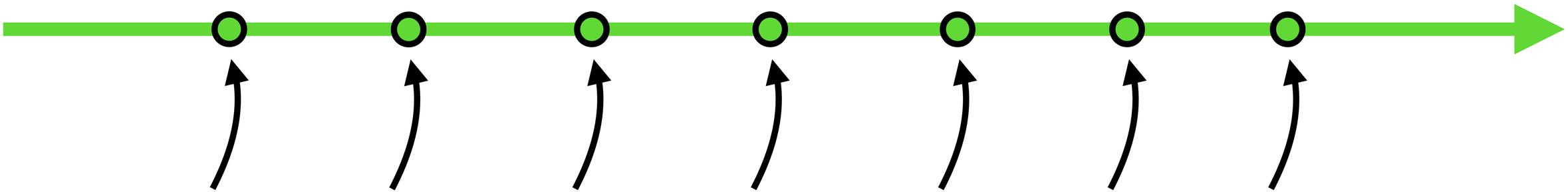
- Far more efficient with tokens
- Make small language models practical

DeepSeek v4 flash is
137x cheaper than Fable
per task

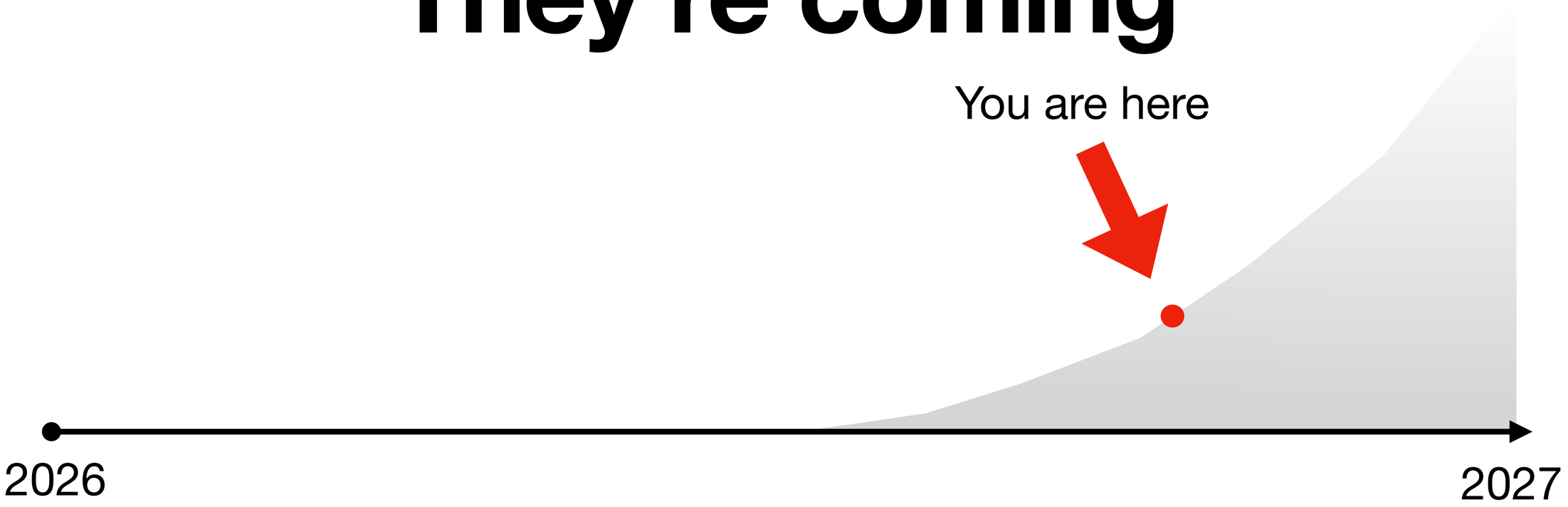


Efficacy

You can eval your own process

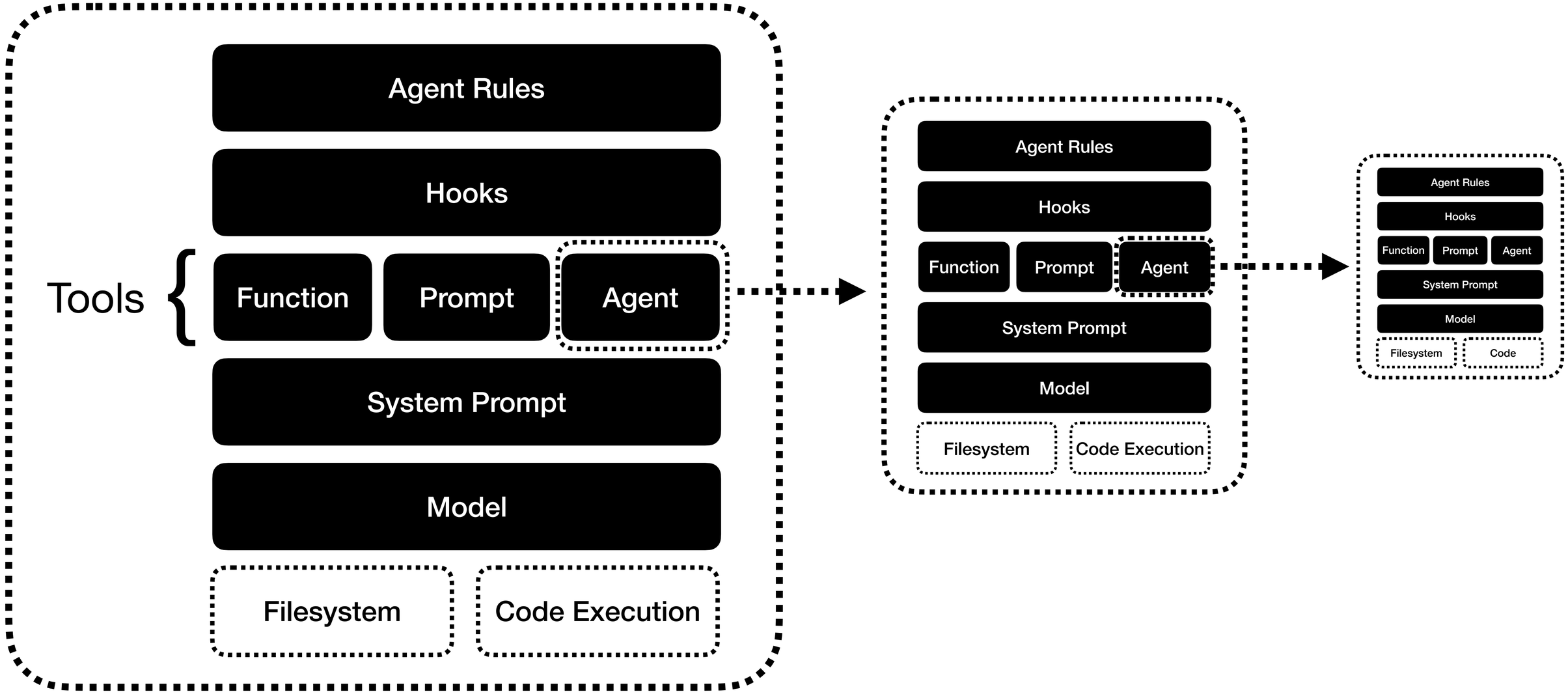


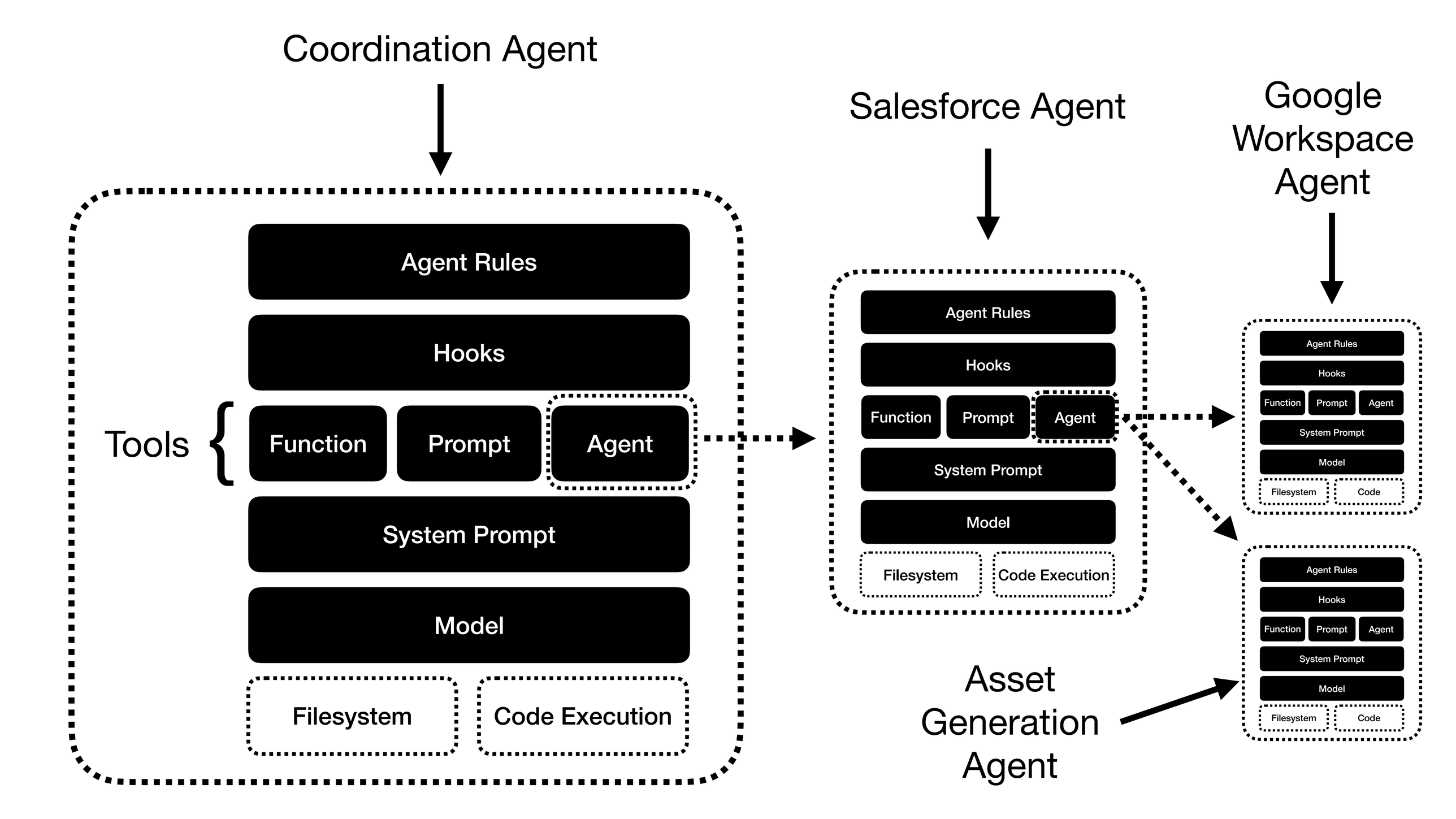
They’re coming

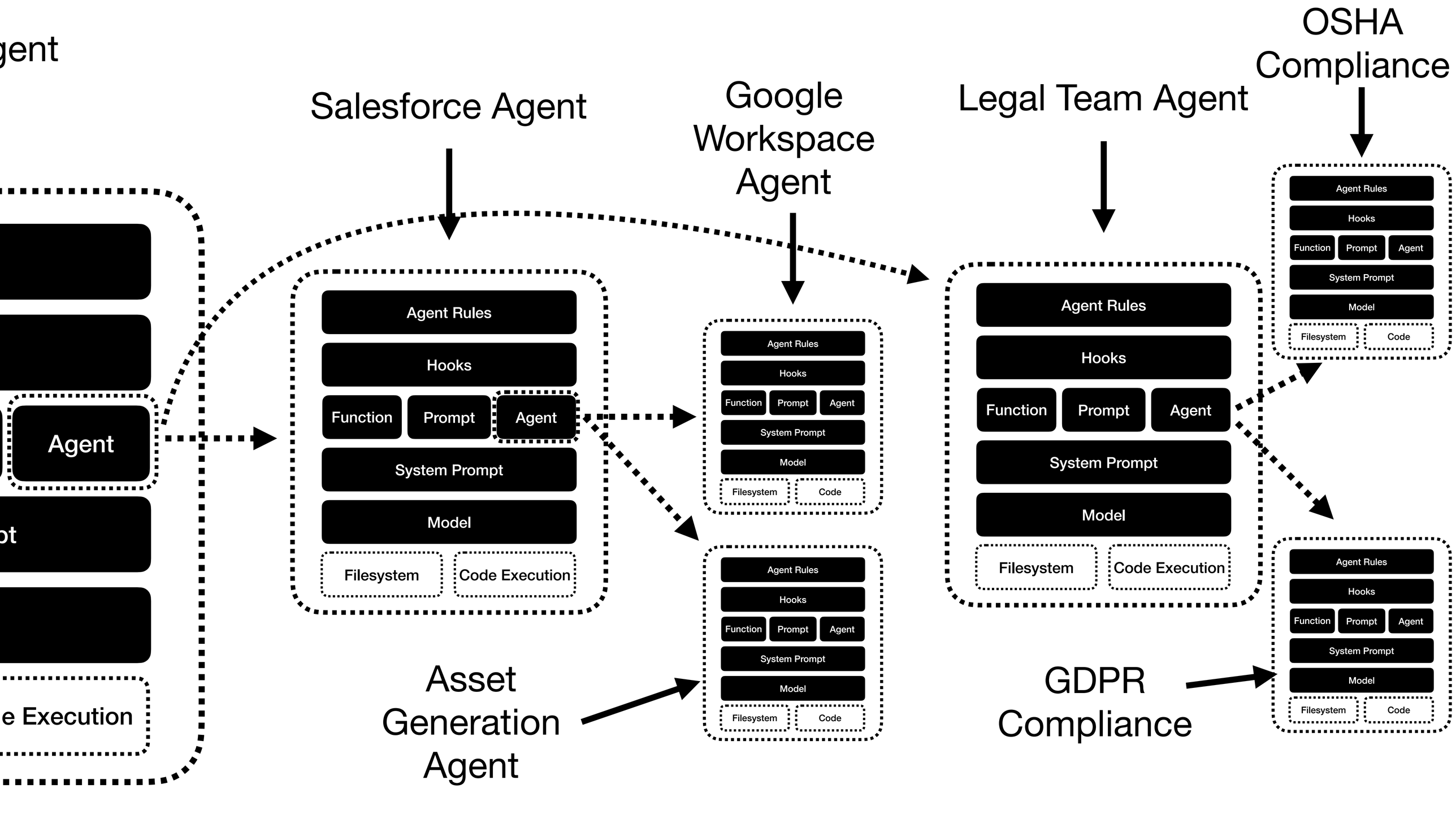


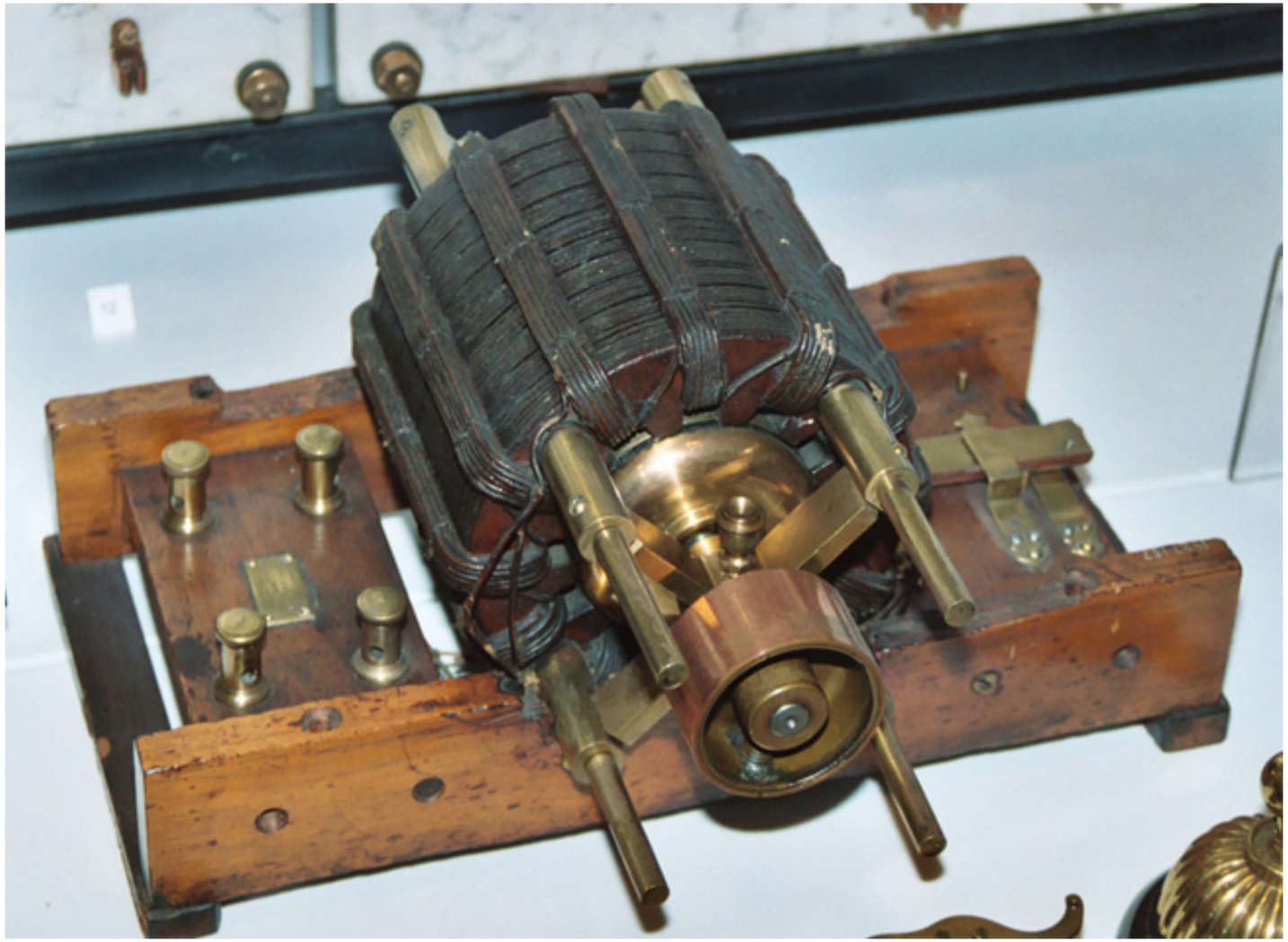
@jpschroeder

What is in an ideal agent?

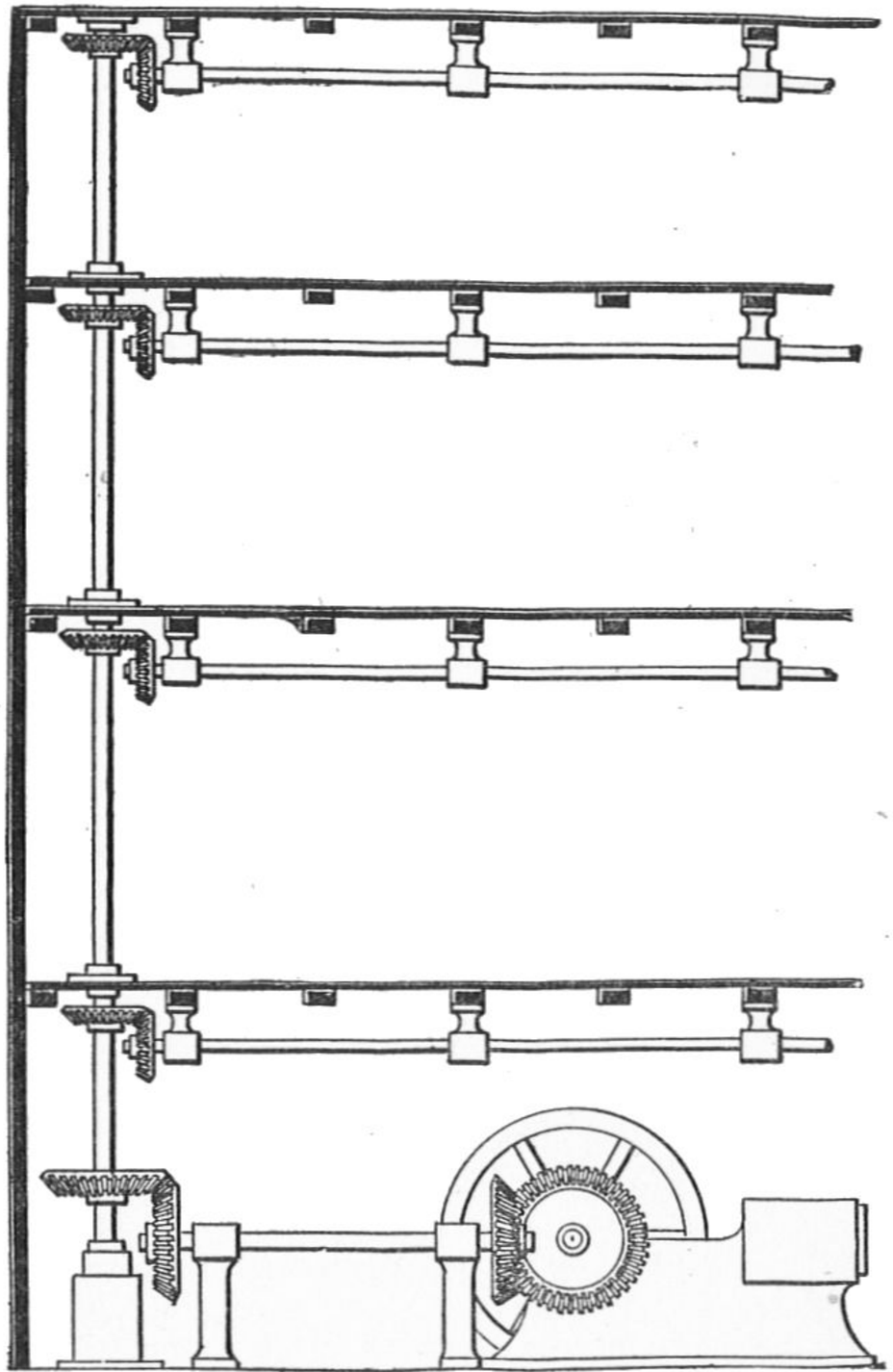



Coordination Agent
Agent Rules
Hooks
Tools
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code Execution
Salesforce Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code Execution
Google Workspace Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code
Asset Generation Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code

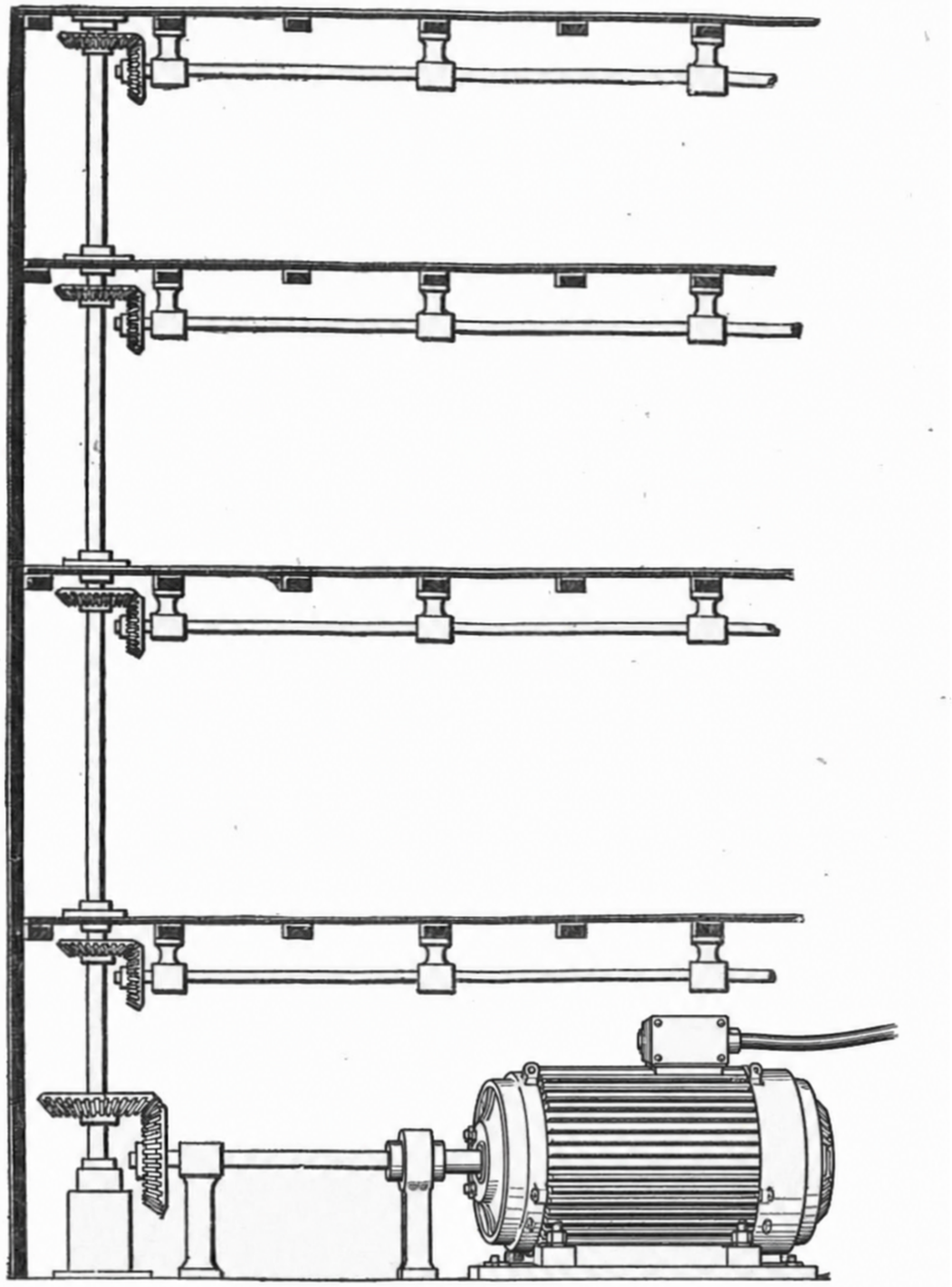

gent
Agent
pt
e Execution
Salesforce Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code Execution
Google Workspace Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code
Asset Generation Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code
Legal Team Agent
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code Execution
OSHA Compliance
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code
GDPR Compliance
Agent Rules
Hooks
Function
Prompt
Agent
System Prompt
Model
Filesystem
Code



@jpschroeder



@jpschroeder



@jpschroeder

Thank you.

StandardAgents

standardagents.ai

info@standardagents.ai



Justin Schroeder
x.com/jpschroeder