

THE AGENTIC SOFTWARE ENGINEERING REVOLUTION

AHMED E. HASSAN
SCHOOL OF COMPUTING
QUEEN'S UNIVERSITY, CANADA



The End of Programming?!!

*“...we can dismiss the notion that computer programmers [sic] will become a powerful elite in the automated corporation. It is far more likely that the **programming occupation will become extinct**...”*



The End of Programming?!!

*“...we can dismiss the notion that computer programmers [sic] will become a powerful elite in the automated corporation. It is far more likely that the **programming occupation will become extinct** ...”*

Herb Simon, founder of CMU-CS Dept, Turing and Nobel winner [Around the appearance of the punchcards in the 1960s]



MindForge

TRAIN HARSHER.
BUILD ENGINEERS,
NOT PARROTS.



ProgramBench

37.98% → 49.51%



DeepSWE

1.76% → 15.92% 9×



973

DENSE TRAJECTORIES

THE TRAINING BROUGHT OUT
THE ENGINEER WITHIN, NOT THE PARROT!!

YO, ADRIAAAAN!





TO ADD IS MACHINE. TO DELETE IS HUMAN.

SENIOR ENGINEERS KNOW WHAT TO REMOVE.

Deletion is engineering judgment.

TODAY'S MODELS

>92%

find the file

<52%

delete the exact line

TEACH THE SKILL

-13.9 pp

incomplete deletion

+5.3 pp

SWE-bench Verified

"To master the code, learn to let go."



BUILDING AN LLM IS LIKE REBUILDING AN AIRPLANE EVERY SIX MONTHS. SOMETIMES, YOU EVEN CHANGE THE ENGINE MID-FLIGHT.

BROWNFIELD MAINTENANCE AT AI SPEED

SOFTWARE → DATAWARE: CODE PATCH → DATA PATCH • REGRESSION TEST → REPEATED EVALUATION • BINARY → CHECKPOINT

THIS IS WHAT SOFTWARE ENGINEERING WAS CREATED FOR.

Rajbahadur · Ebrahimi · Chen · Hassan | [arXiv:2608.31102](https://arxiv.org/abs/2608.31102)



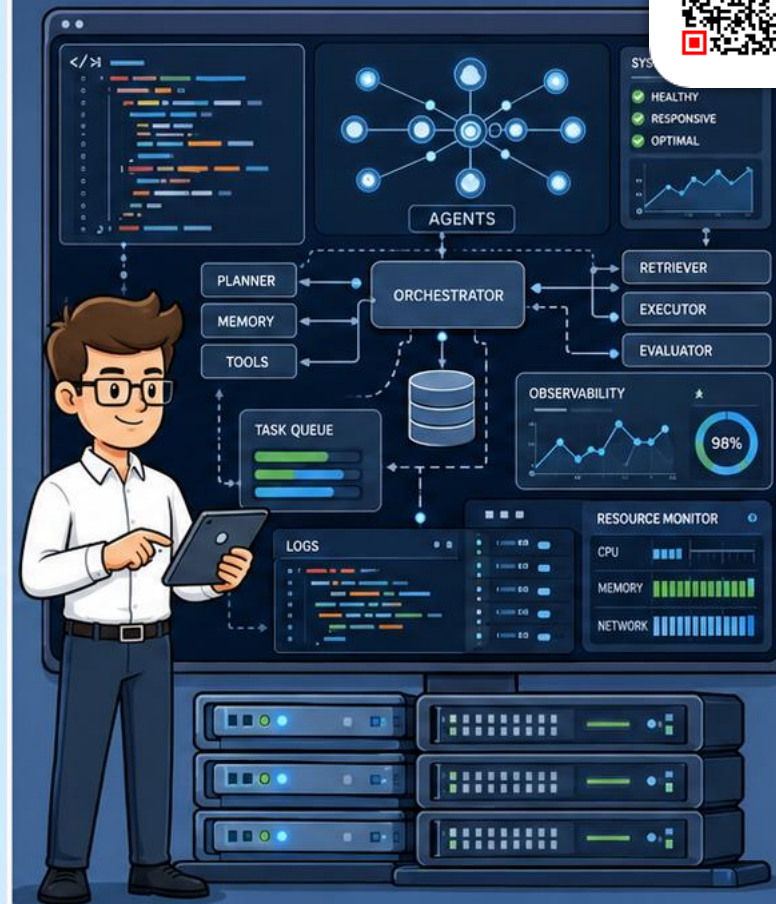
VIBE CODING



VIBE CODING



VIBE ENGINEERING



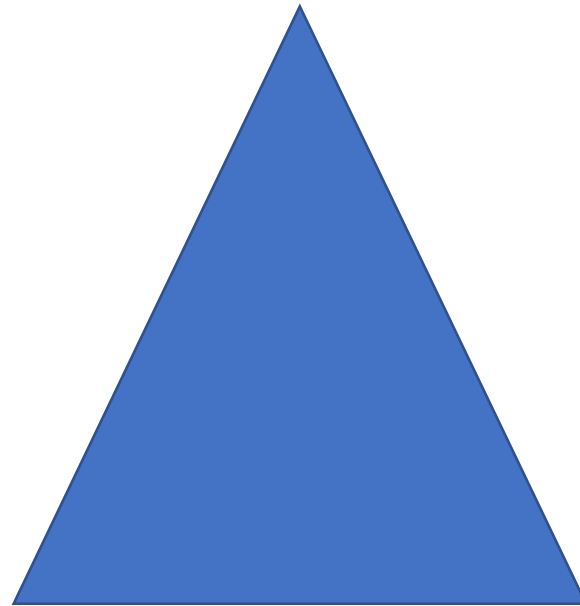
AGENTIC SE



Software Engineering is Changing!

**AI-Native
Software**

~~Product/Value~~



**AI-Native Process
(Vibe Coding,
Vibe Engineering,
Agentic SE)**

~~Process~~

~~Actors~~

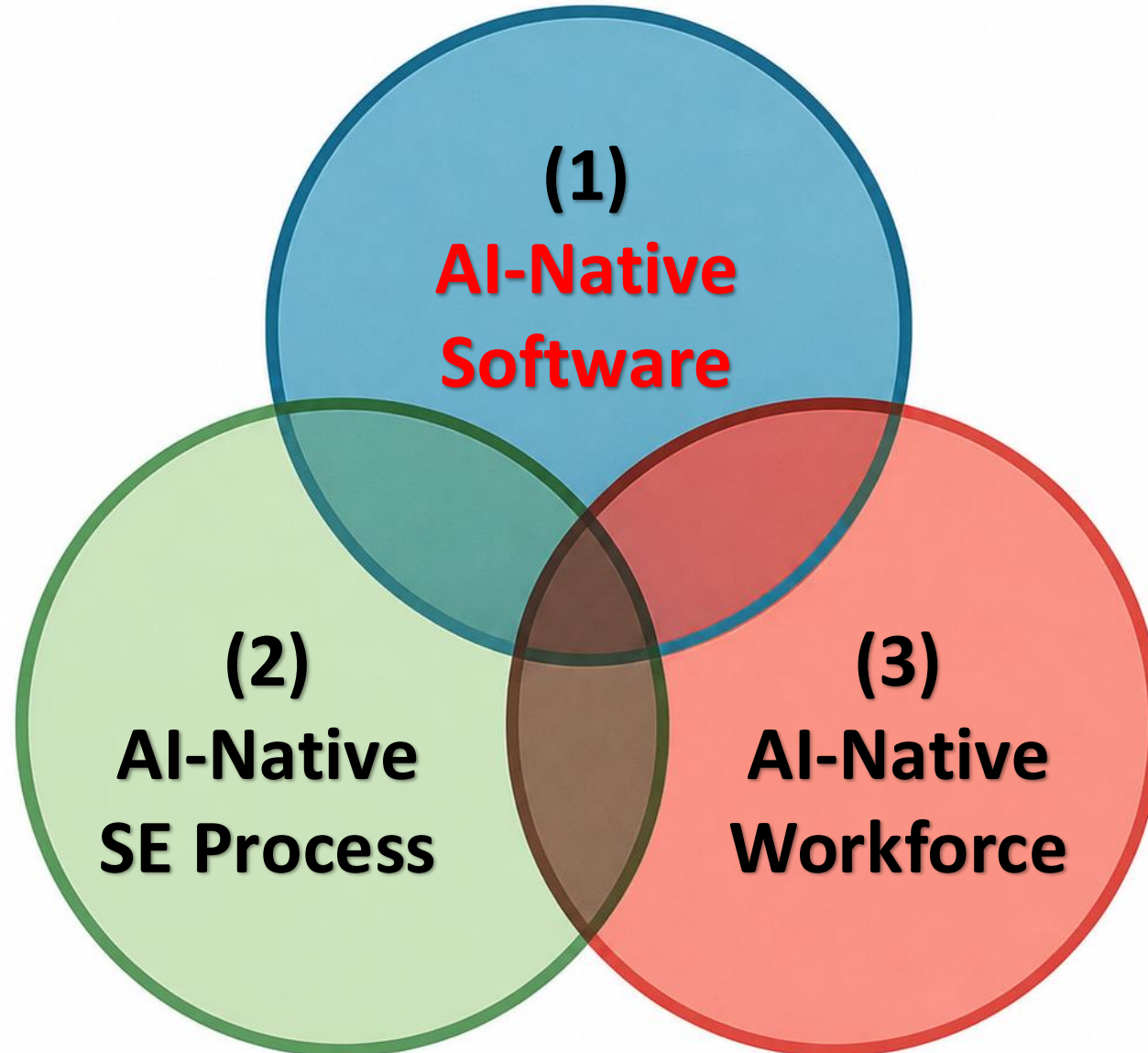
**Expert Humans,
Non-Expert Humans,
AI Agents**



**Unfortunately,
this is where we
are today!**



AIware: Software for All by All!

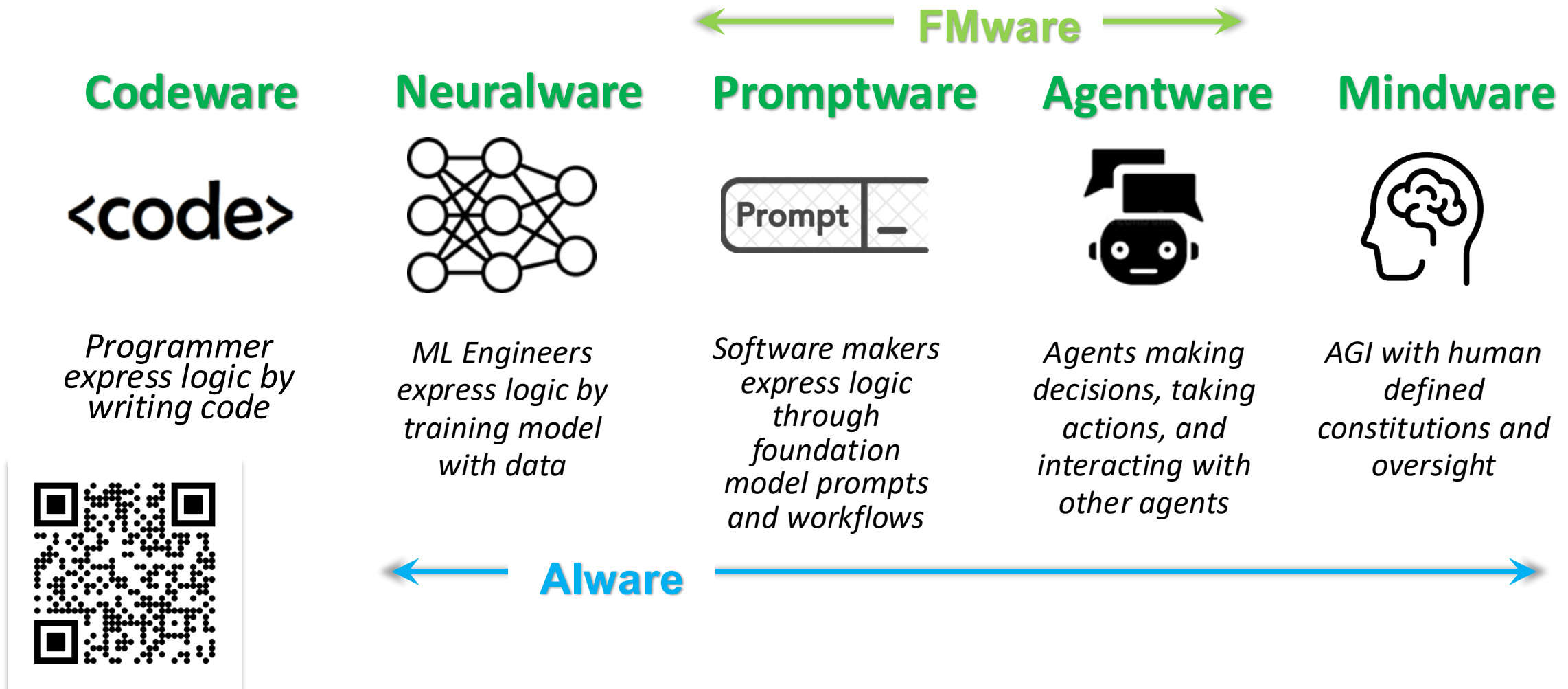


+
(4)
AI-Native
SE Pedagogy

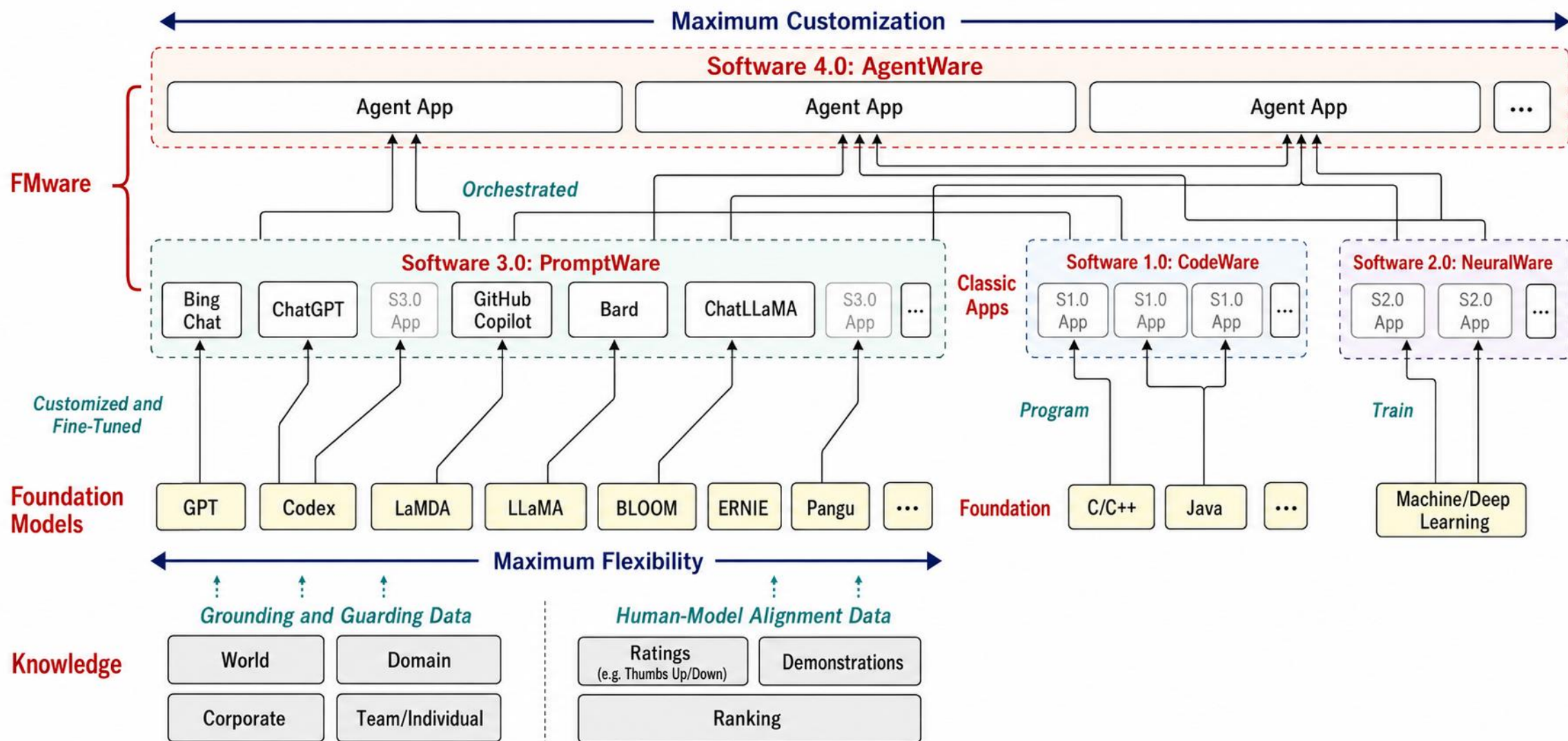
<https://www.aiwareconf.org/>
<https://www.aiwarebootcamp.io/>

AI-Native Software:

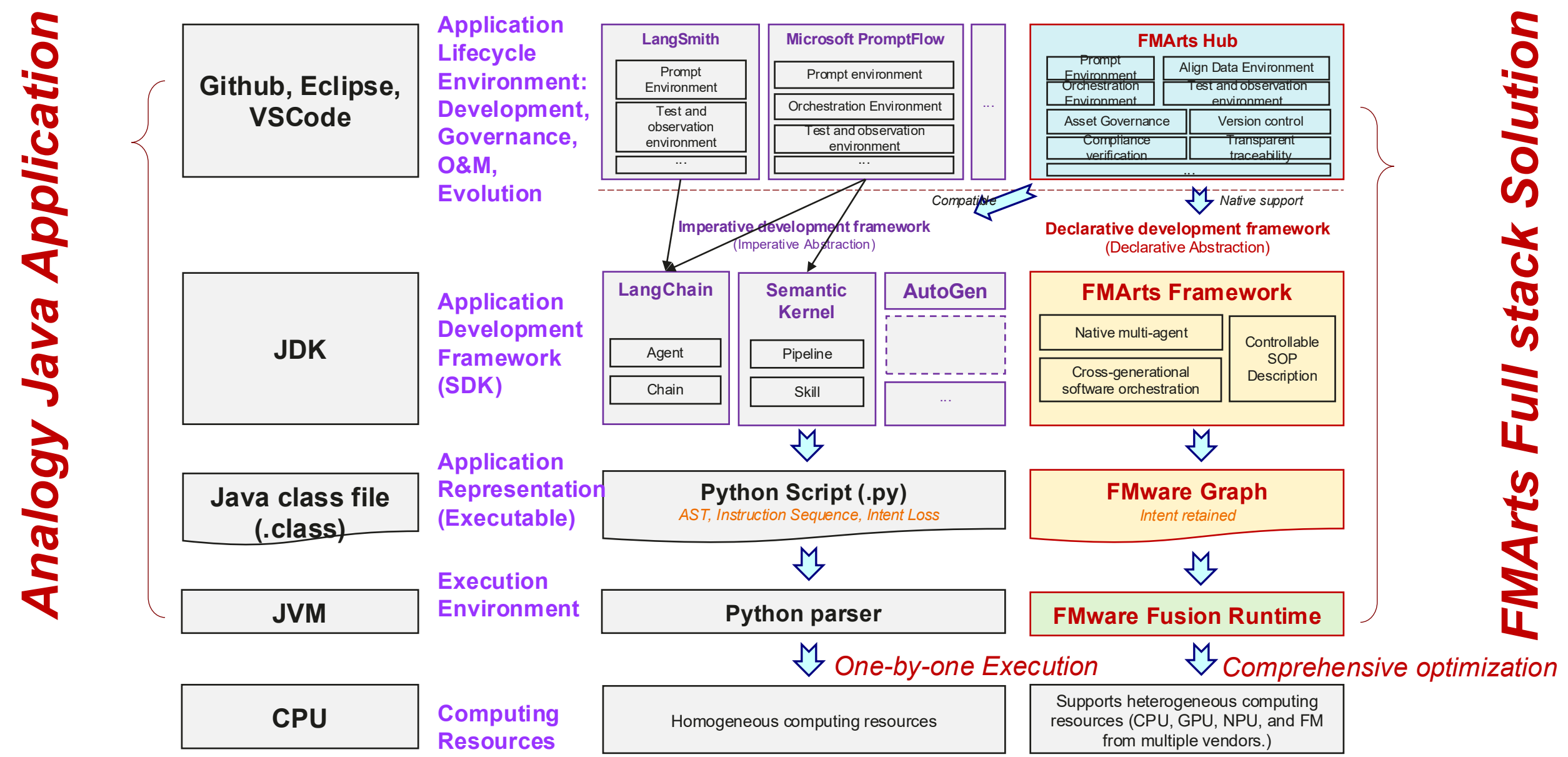
The Evolution of Software Generations each with a new form, lifecycle, managed assets, and roles (aka Engineering Paradigm)



Future Software is a Composition of Many Generations of Software and FMs



We must rethink the engineering stack for FMware



We are just starting... Tackling these challenges requires technology breakthroughs, standards, and community efforts

C1: Managing alignment data

C7: Operation & semantic observability

C2: Crafting effective prompt

C8: Performance engineering

C3: Multi-generation software

C9: Testing under non-determinism

C4: Degree of controllability

C10: Siloed tooling & lack of process

C5: Compliance & regulations

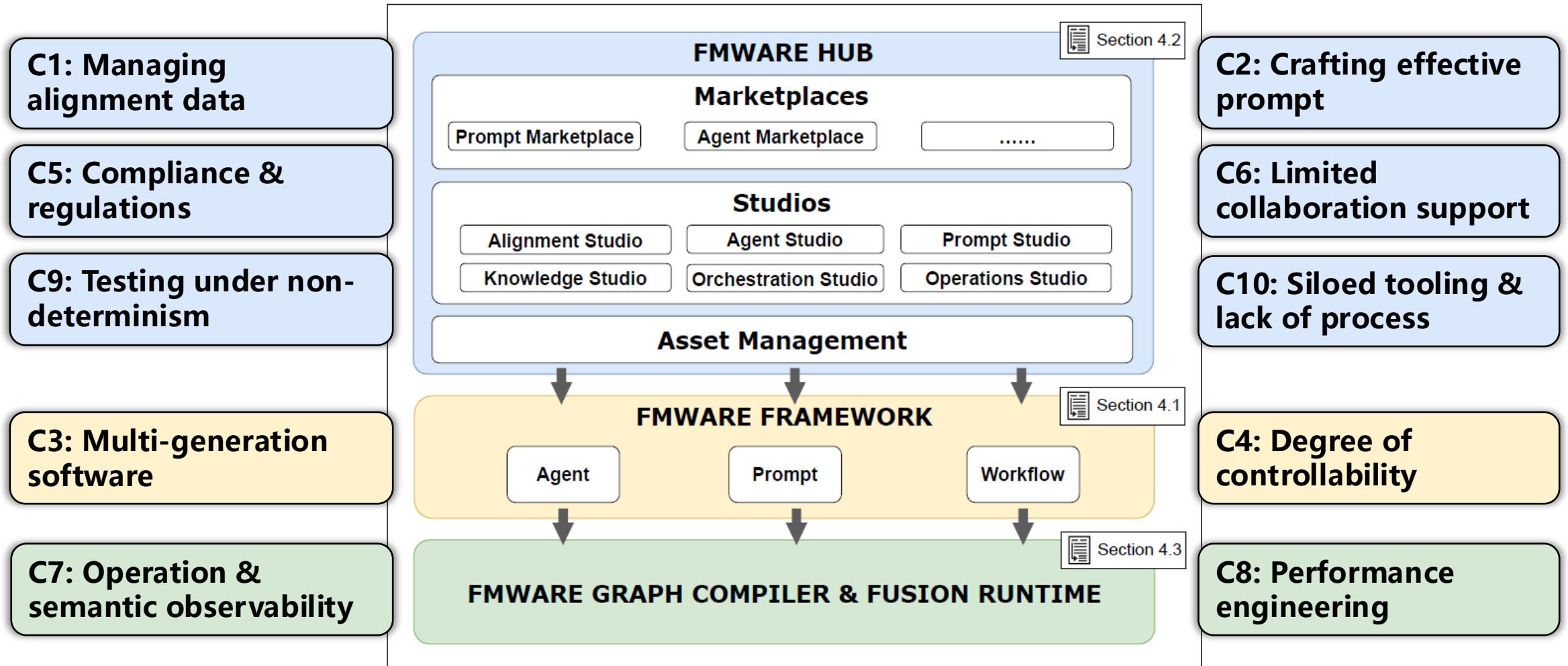
C6: Limited collaboration support

More details at:
<https://arxiv.org/abs/2402.15943>

<https://aiwareconf.org>
<https://fmse.io/>
<https://se4ai.org/>



FMArts: Our FMware lifecycle engineering stack



The World's First Prompt Debugger



(4) Model Config

Model Config

Model

llama-2-test Select

repetition_penalty 1.2

top_k 50

max_new_tokens 300

temperature

Design

Load Save As Save

Structured Prompt Chain

Please specify input variables using the `{{SOME_VARIABLE}}` syntax. When using the `OUTPUT FORMAT` component, please also add the `POST CONDITION` component for each `[[MASK]]` in that `OUTPUT FORMAT`.

Memory UNDER DEVELOPMENT

(1) Prompt Editor

Instructions ##Instruction: \n Save Save As Load

Translate English to French

Debug Run

Inputs

Output

(2) Output

-> French: Combien t'es-tu?

Prompt Used

##Instruction: Translate English to French

Prompt Interpretation

#	#	Instruction	:	Translate	English	to	French	<0x0A>	<0x0A>	#	#	Examples	word
0.02	0.02	0.06	0.03	0.05	0.03	0.03	0.03	0.03	0.03	0.02	0.02	0.05	
:	<0x0A>	<0x0A>	English	:	Hello	<0x0A>	<0x0A>						
0.03	0.03	0.03	0.03	0.03	0.03	0.03	0.03						

Token Importance

Low High

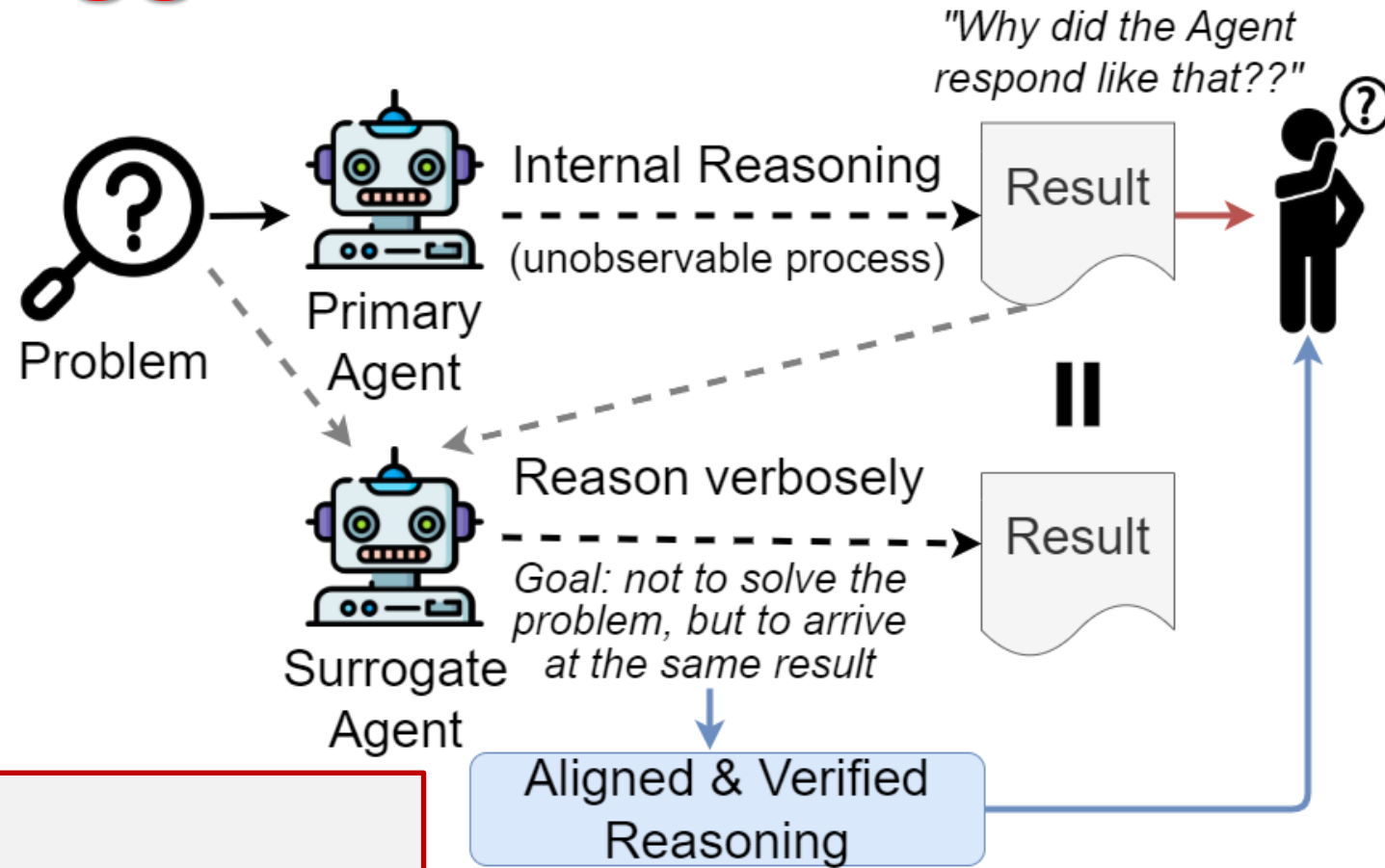
(3) Prompt Explanation (e.g., word-level explanation)



Watson: The World's first Agent Debugger



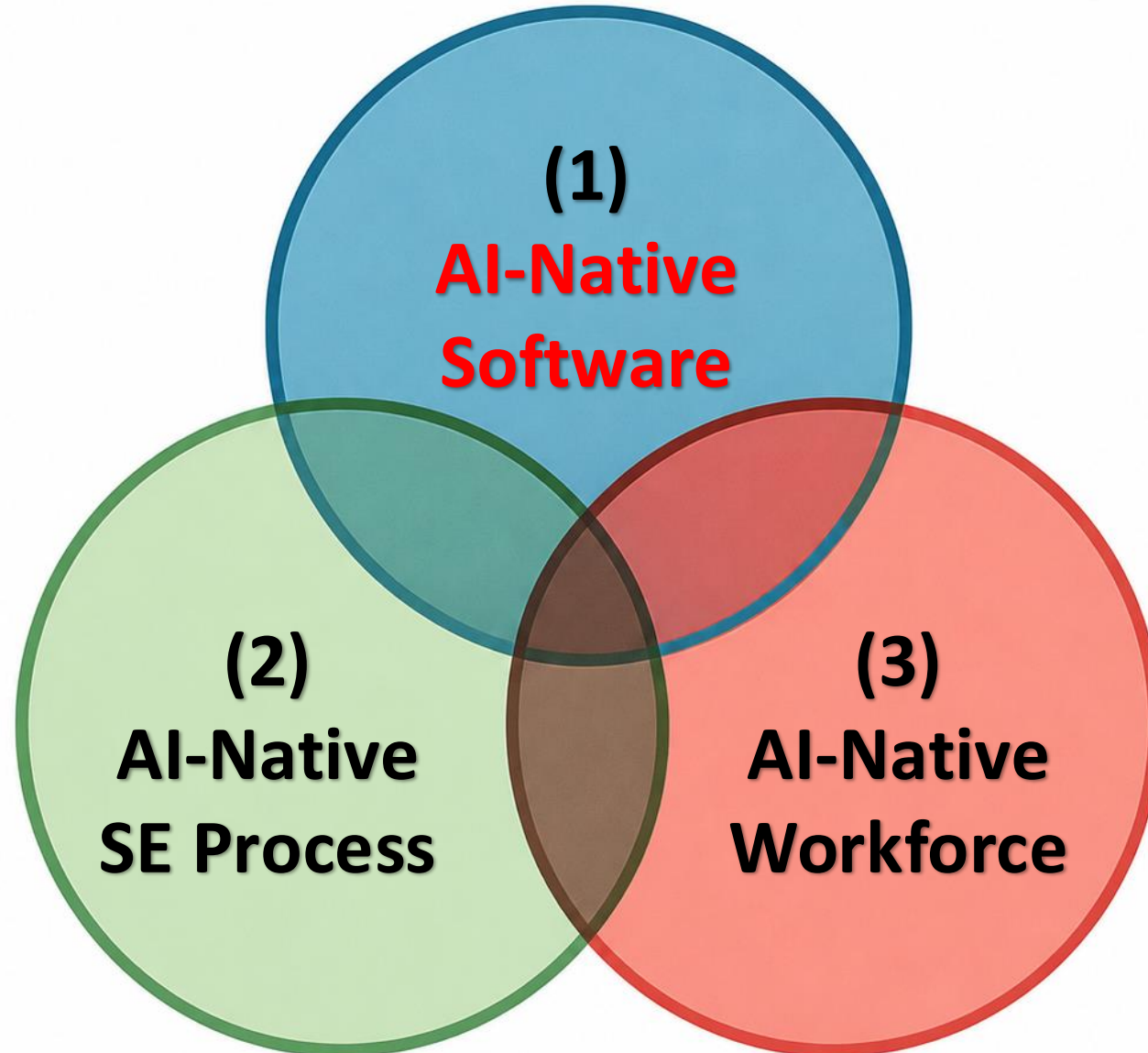
- Capture agent's internal reasoning without interfering with behavior
- Ensure reasoning is aligned & verified with agent's focus



- ✓ **Fix Errors** – Spot & guide reasoning
- ✓ **Boost Insight** – Help devs & agents learn
- ✓ **Improve Together** – Enable collaborative reasoning



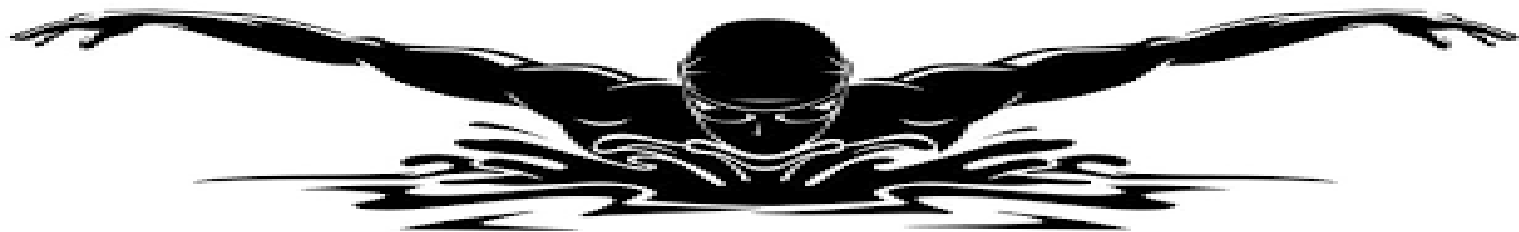
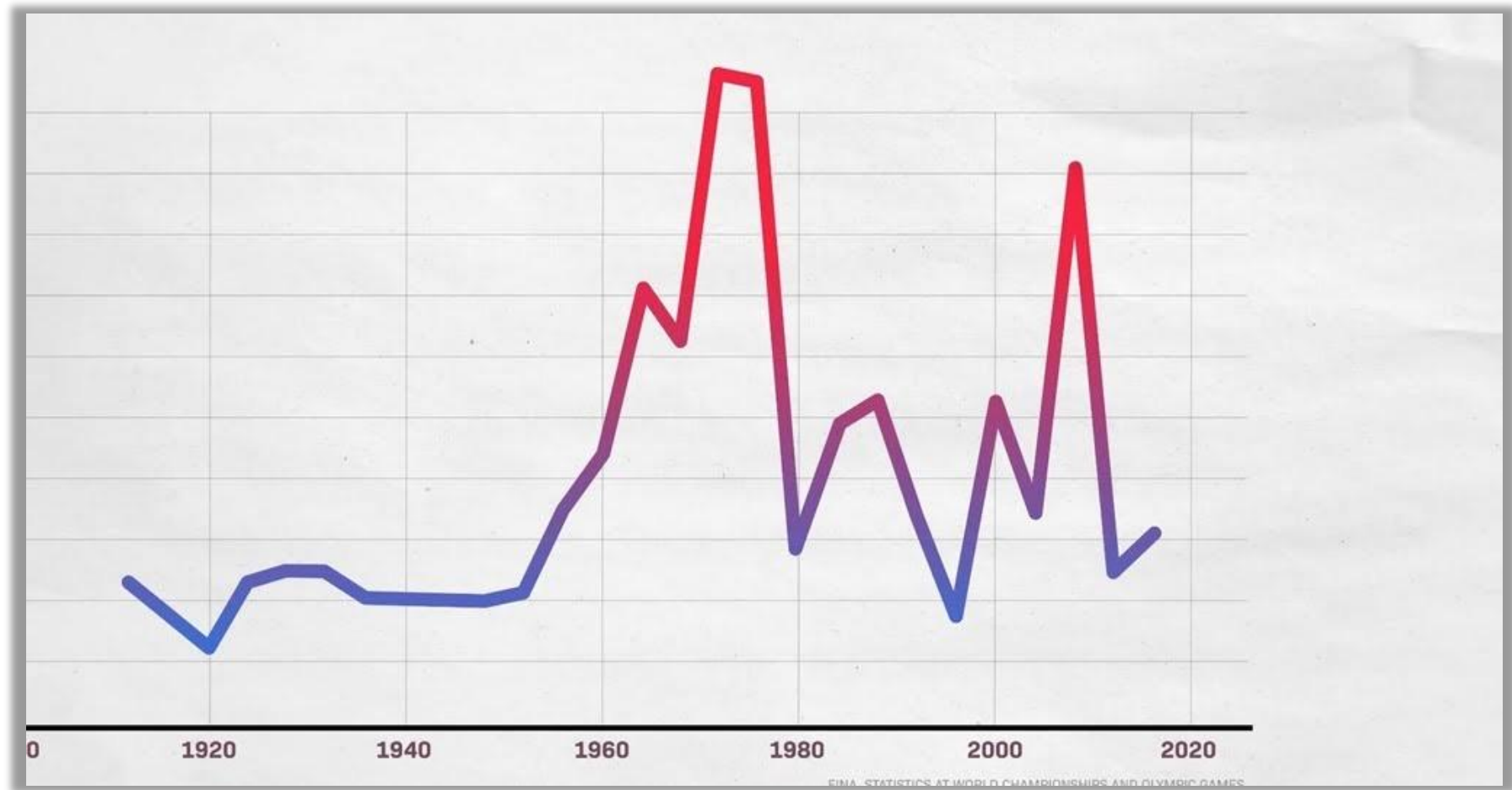
AIware: Software for All by All!



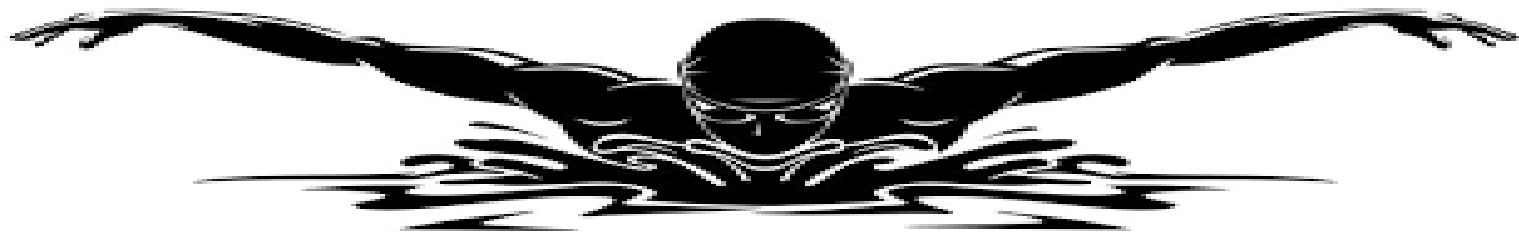
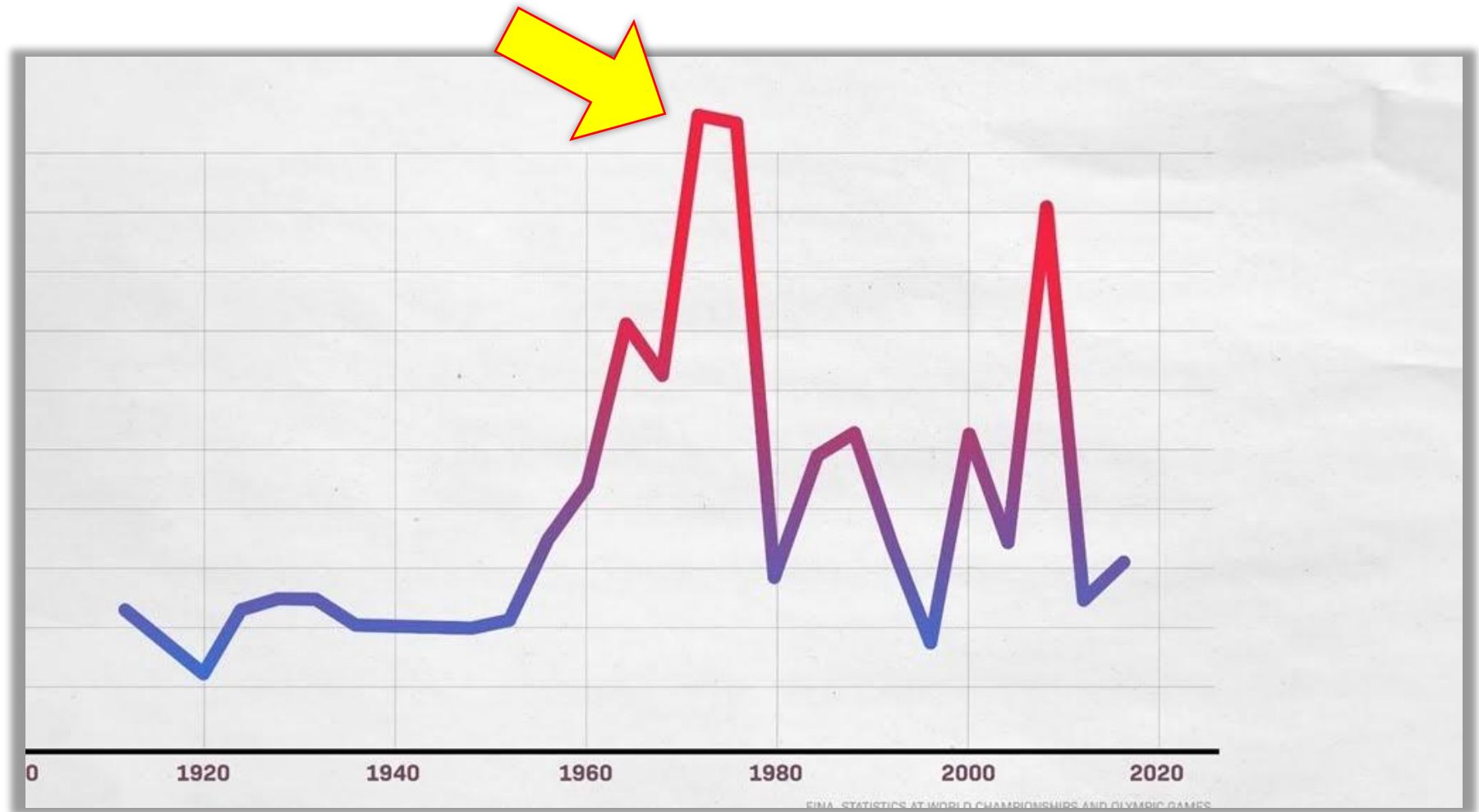
+
(4)
AI-Native
SE Pedagogy

<https://www.aiwareconf.org/>
<https://www.aiwarebootcamp.io/>

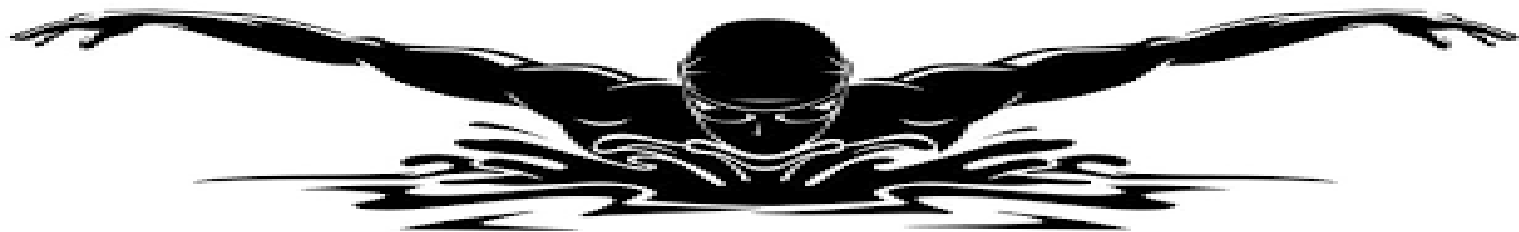
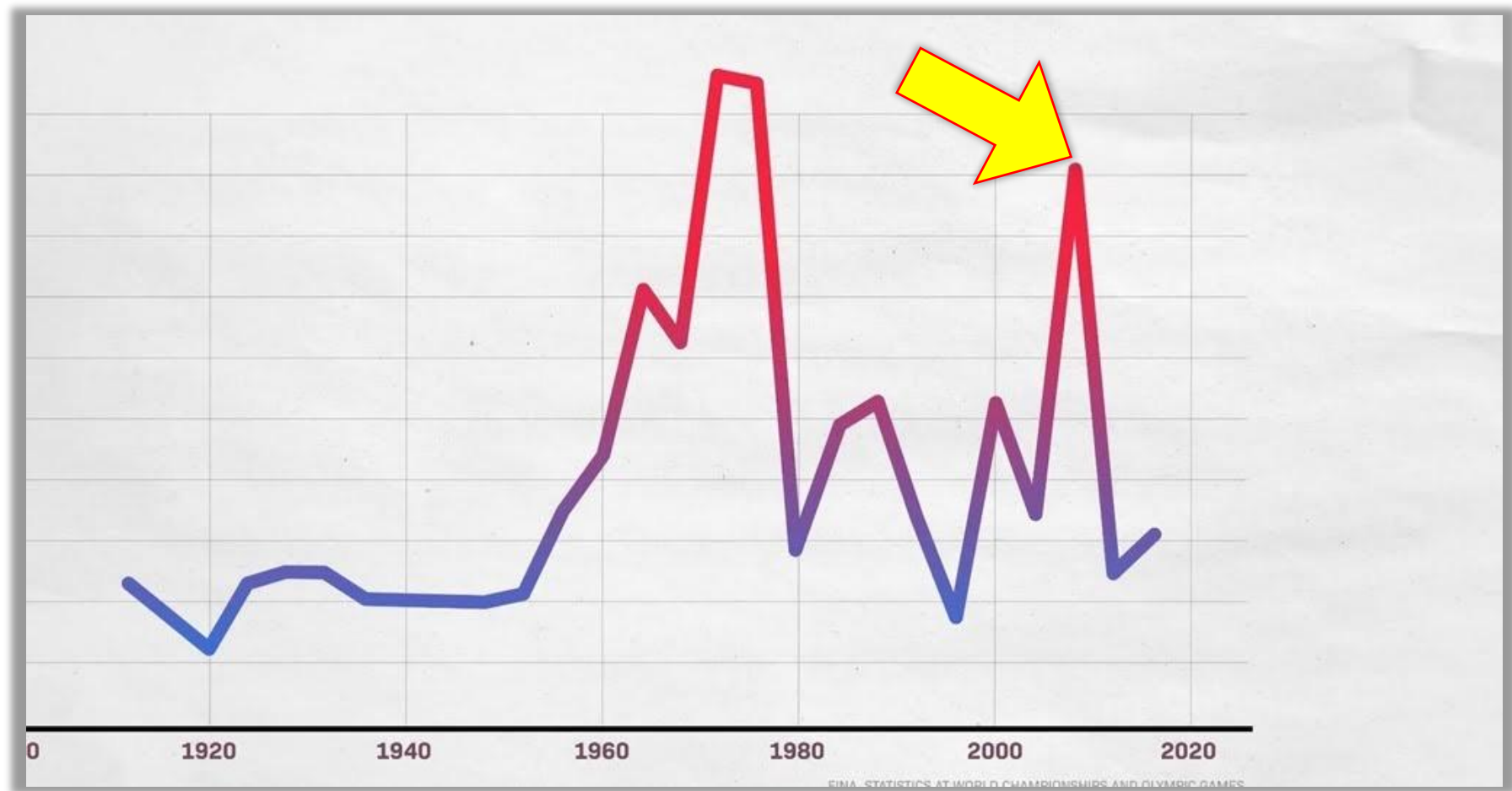
**# of
broken records
per year
in Olympics
& World
championships
for swimming**



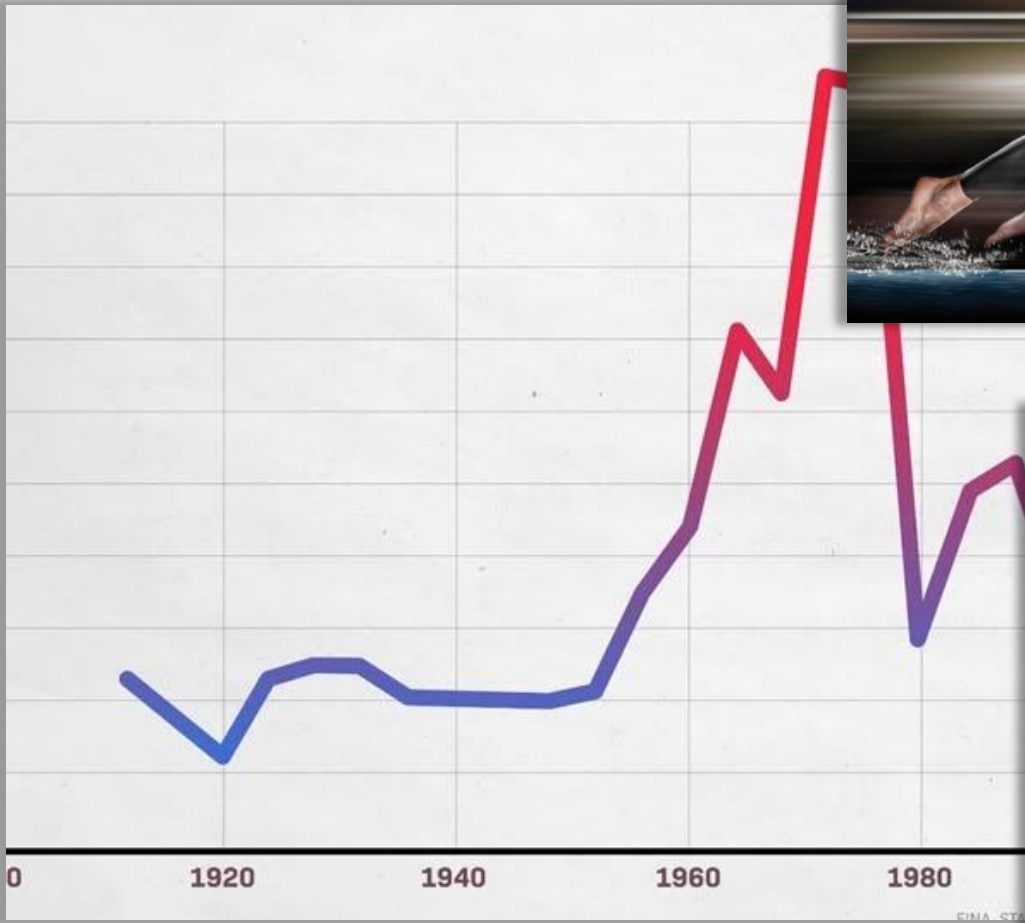
**# of
broken records
per year
in Olympics
& World
championships
for swimming**



**# of
broken records
per year
in Olympics
& World
championships
for swimming**



of broken records per year in Olympics & World championships for swimming



Built for speed

The LZR RACER, a new swimsuit developed by Speedo with the help of NASA, will be worn by swimmers in the 2008 Summer Olympics.

Construction
More than 100 body scans of swimmers done to study shapes for ideal suit

- 1. **Bonded seams**
 - First fully bonded bodysuit; ultrasonically welded seams
 - Zipper bonded into suit for smooth finish
- 2. **Support**
 - Core stabilizer allows for best body position
 - Panels deliver optimum streamlined shape
- 3. **Fabric**
 - Unique, lightweight, water-repellent fabric
 - Powerful compression reduces drag

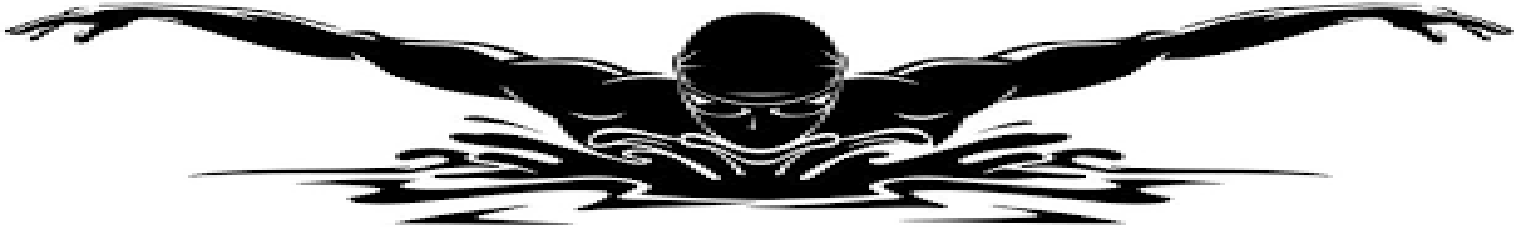
Rocket in the water
Main focus was to reduce friction or drag a swimmer experiences in water

- **Hydrodynamic drag** -----
Water passing over the body as it moves through the pool
- **Form drag** ○
Bumps, curves and muscle oscillations on swimmer's body; hinder progress in water
- **Goal** make suit as smooth as possible

Water easily passes over fabric

Panels hold steady and firm as swimmer moves through water

© 2008 MCT
Source: Speedo
Graphic: Marina Vingling



The “Dick Fosbury Flop” in High Jump



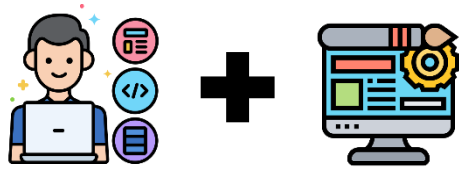
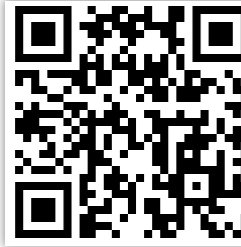
The Straddle



Fosbury Flop

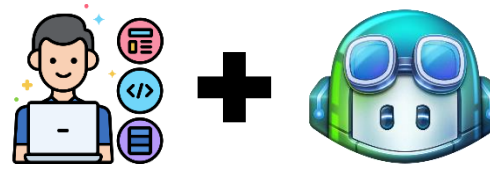
Mexico 1968 Olympics

AIware: Rethinking Software Engineering!



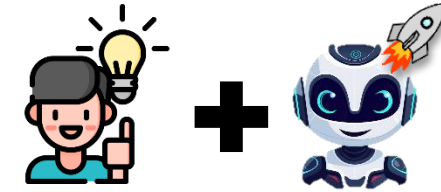
Software Engineering 1.0

- **Code-first**
- **Tools** supporting **traditional** SE Process activities
- Powered by **program analysis technologies**



Software Engineering 2.0

- **Code-first**
- **AI models** supporting **traditional** SE process activities (AI4SE)
- Powered by **expensive data-driven models** with limited capabilities



Software Engineering 3.0

- **Intent-first**, conversation-oriented development
- **AI-native** SE process maximizing the strengths of human (requirements) & AI (implementation)
- Powered by **efficient knowledge-driven** models with advanced reasoning capabilities



The Brain's last stand



“We humans are trying to figure out our next move” Dan Rather, NBC

“A **weak** human player + machine
+ **better process**

is superior

to a very powerful machine alone,

But more remarkably is superior to

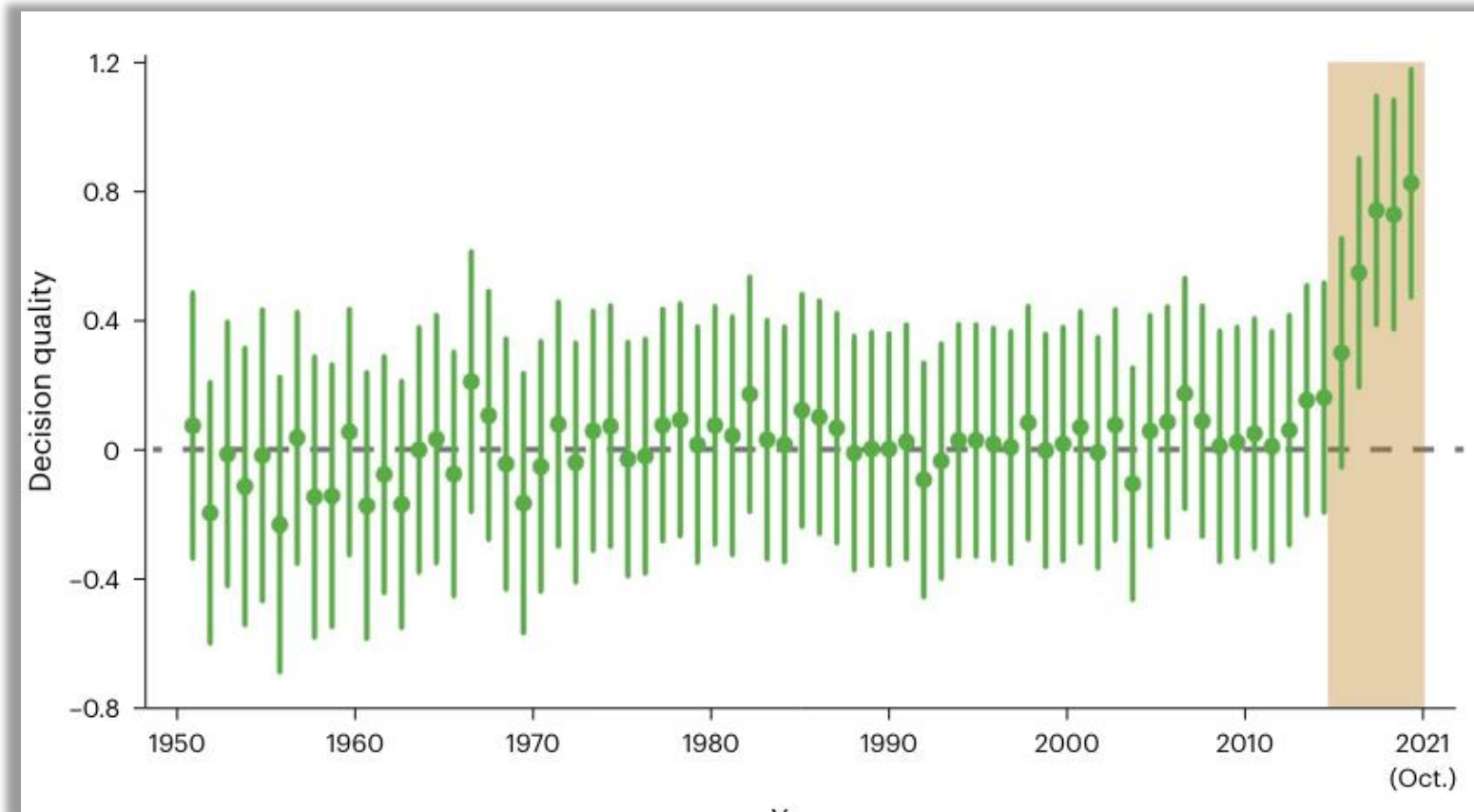
a **stronger** human player + machine
+ **inferior process**”

Gary Kasparov



AI can make developers better

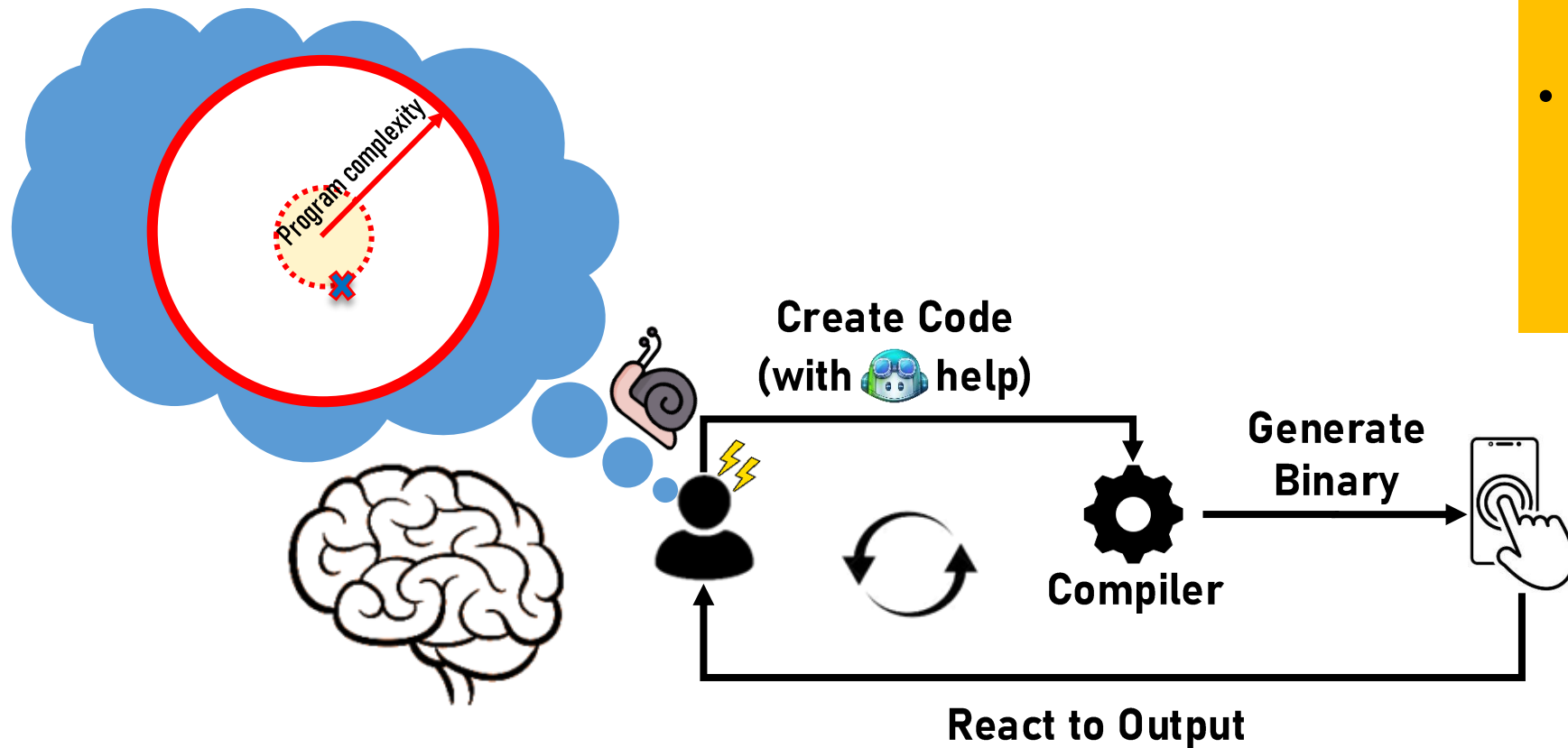
- Strong teams attract strong candidates, making them stronger and keeping them longer → **Your AI teammate is a coach that is making you a better developer**



Decision quality of professional Go players evaluated through an algorithm that is performing at a superhuman level. Decision quality significantly increased after Lee Sedol was beaten by AlphaGo on 15 March 2016 (shaded area) [Brinkmann 2023]

Programming 2.0 [AI Assisted]

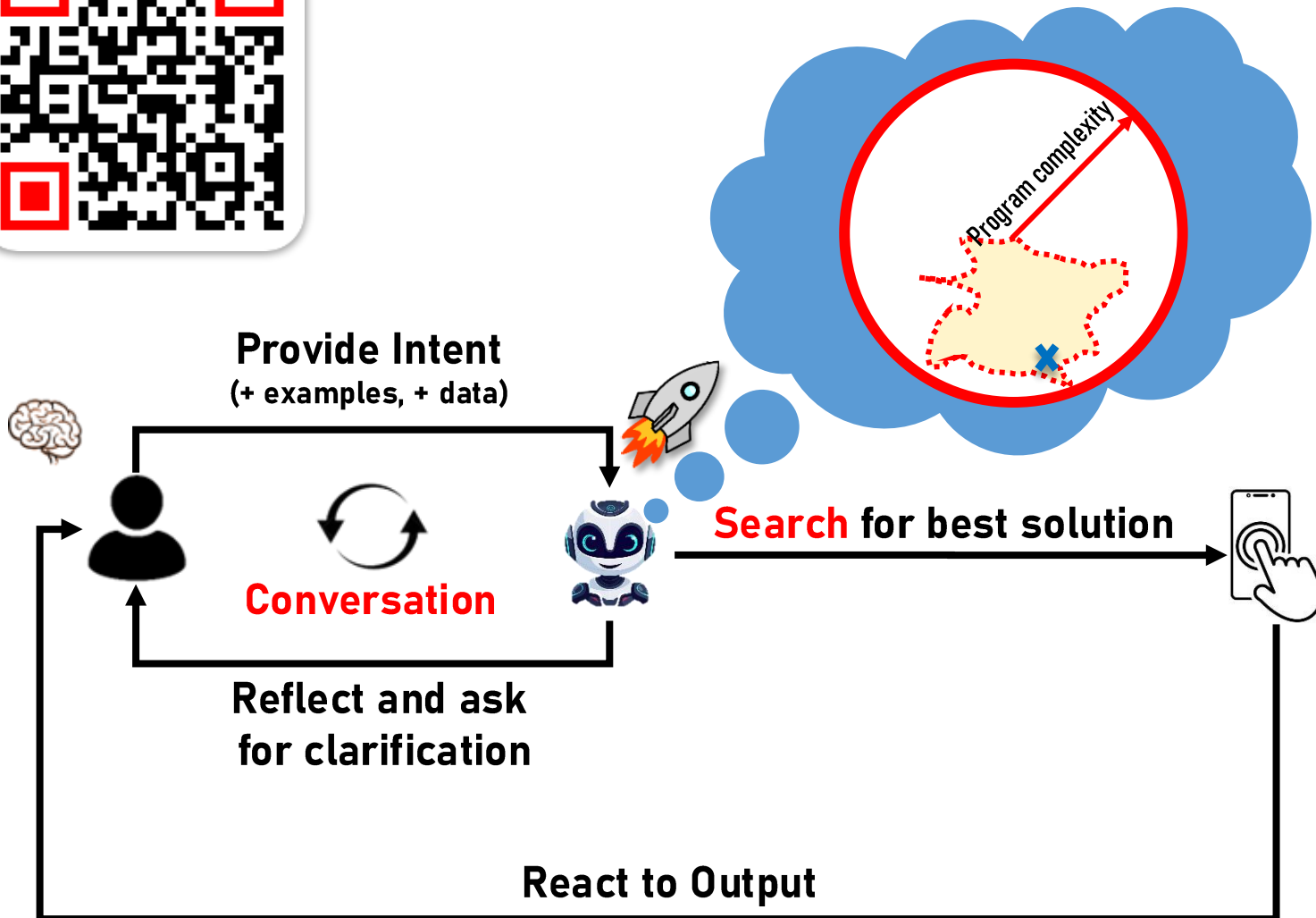
(Write Code with AI help then Compile)



- Human drives the code creation process loop
- The code complexity is limited by human's ability to express *and maintain* a complex problem in code

Programming 3.0 [AI-Native Compiler.NEXT]

(Define Search Space, Give Data, AI Searches to create code)



- Human and AI share clearly aligned goals
- AI drives the end-to-end code creation loop
- Code complexity is unconstrained as the AI synthesizer searches for the software solution that maximizes a fitness function aligned with those goals
- Realization happens in Alware or Codeware
- The loop can be restarted at any time as long as we archive the human–AI conversations

POSTCARDS FROM THE TRENCHES

~~"I haven't coded in months."~~

TRUTH: Engineers must "scrub in" for hard cases to maintain judgment, avoiding intuition decay.

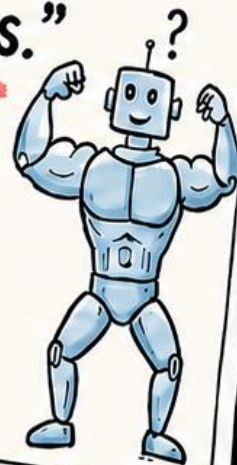


The Tequila-on-the-Beach Fantasy.

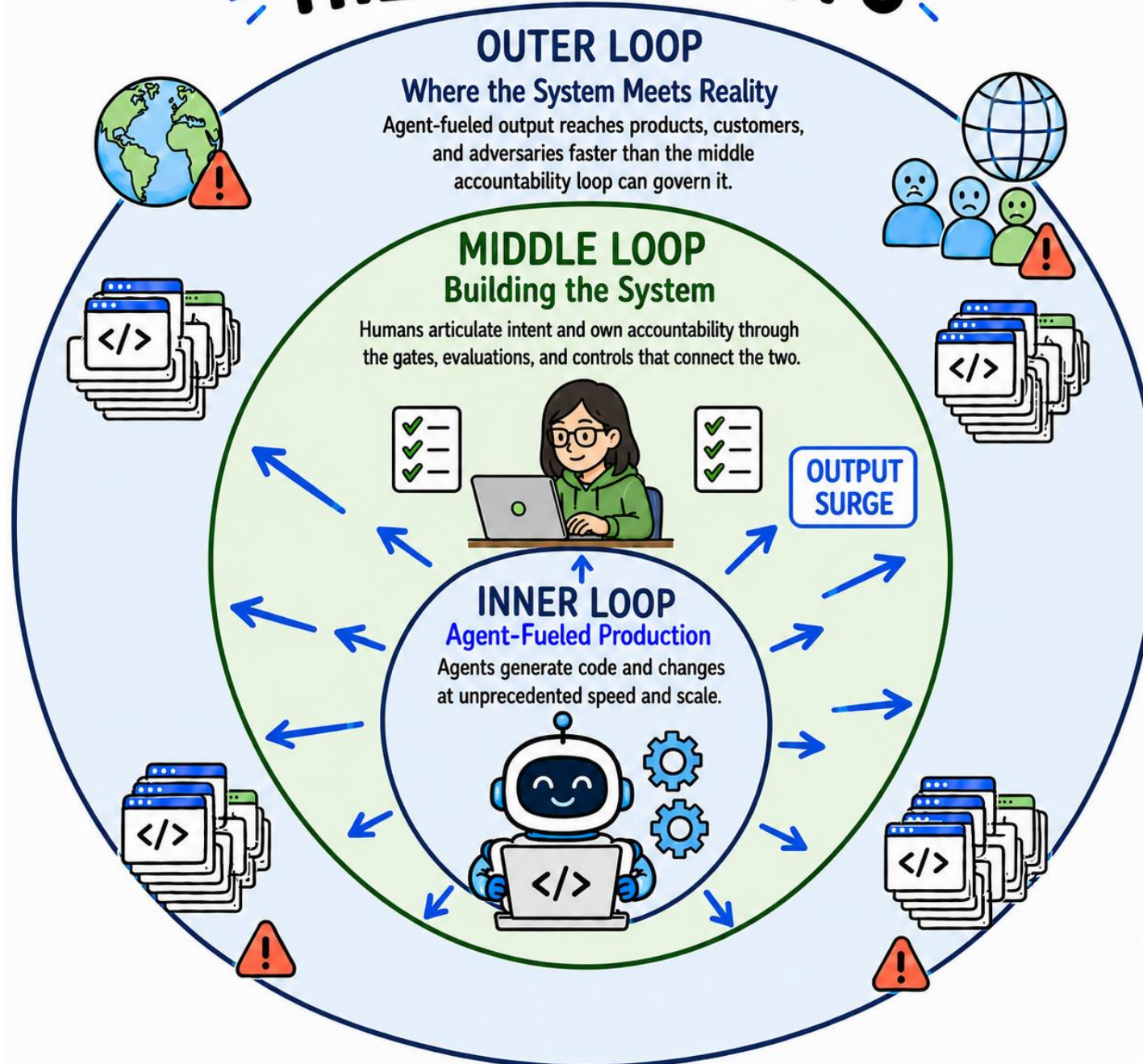
TRUTH: High-leverage agents require "always-on approvers," leading to micromanagement & Trust Debt.

~~"Agents are strong Engineers."~~

TRUTH: Agents are strong hackers but poor engineers, lacking scar tissue, theory, and long-horizon thinking.

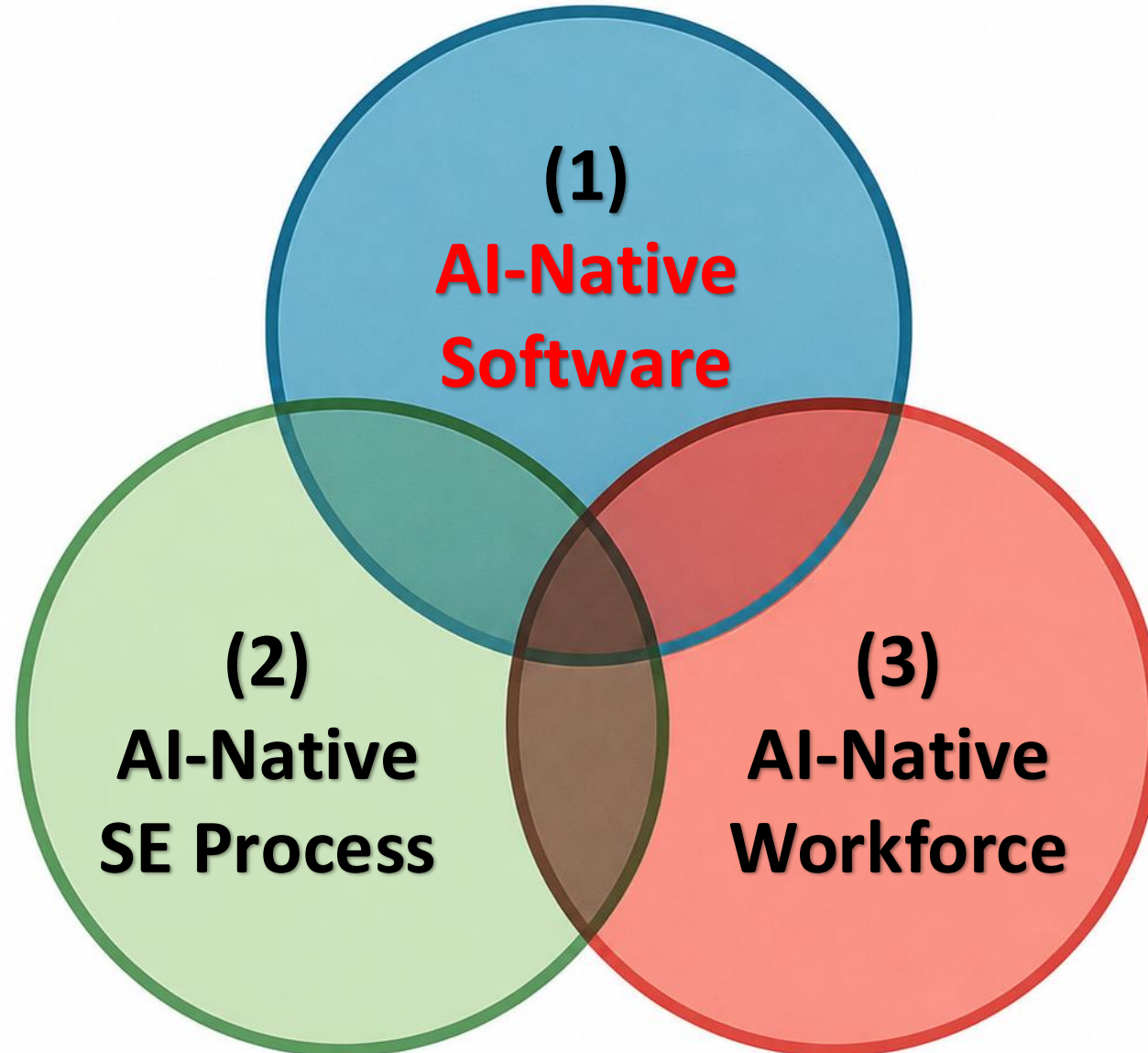


THE THREE LOOPS





AIware: Software for All by All!



+
(4)
AI-Native
SE Pedagogy

<https://www.aiwareconf.org/>
<https://www.aiwarebootcamp.io/>

The Duality of Software Engineering

SE for Humans

**Agent Command
Environment
(ACE)**

SE for Agents

**Agent Execution
Environment
(AEE)**

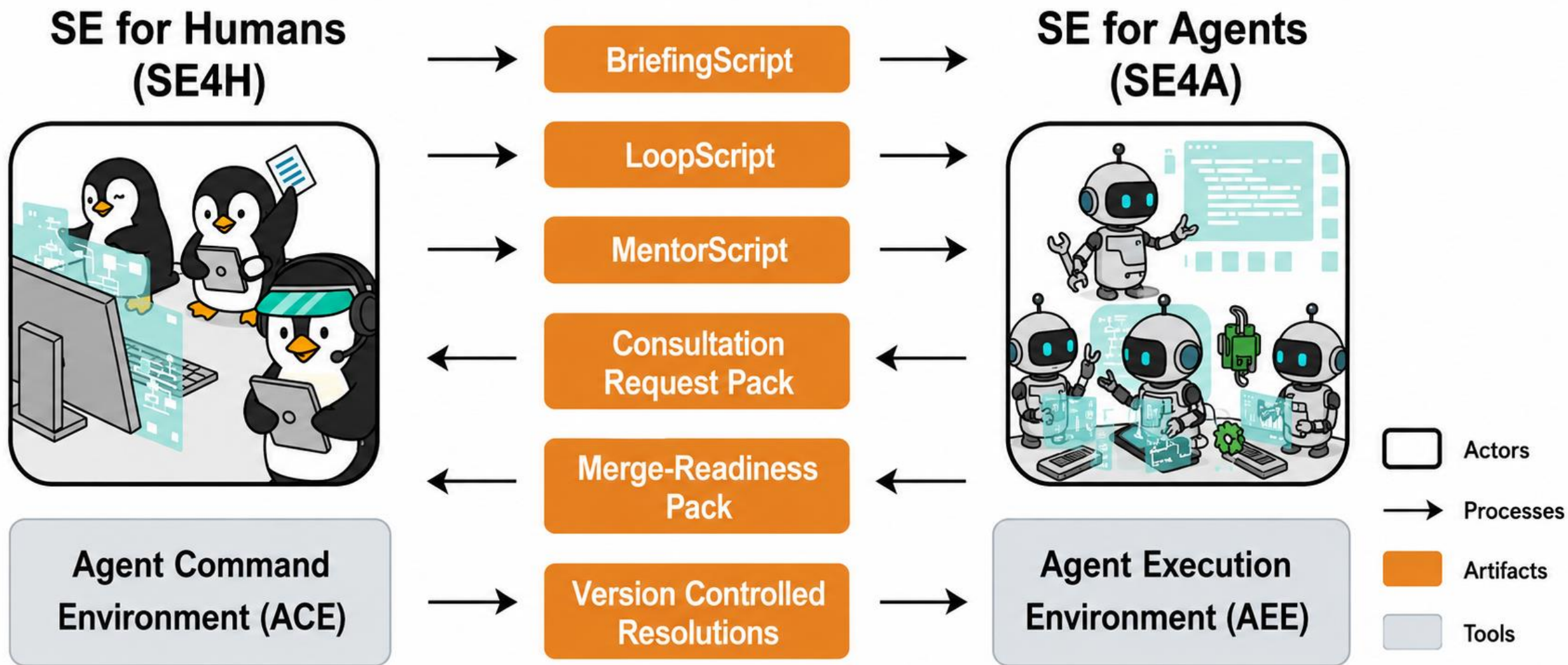
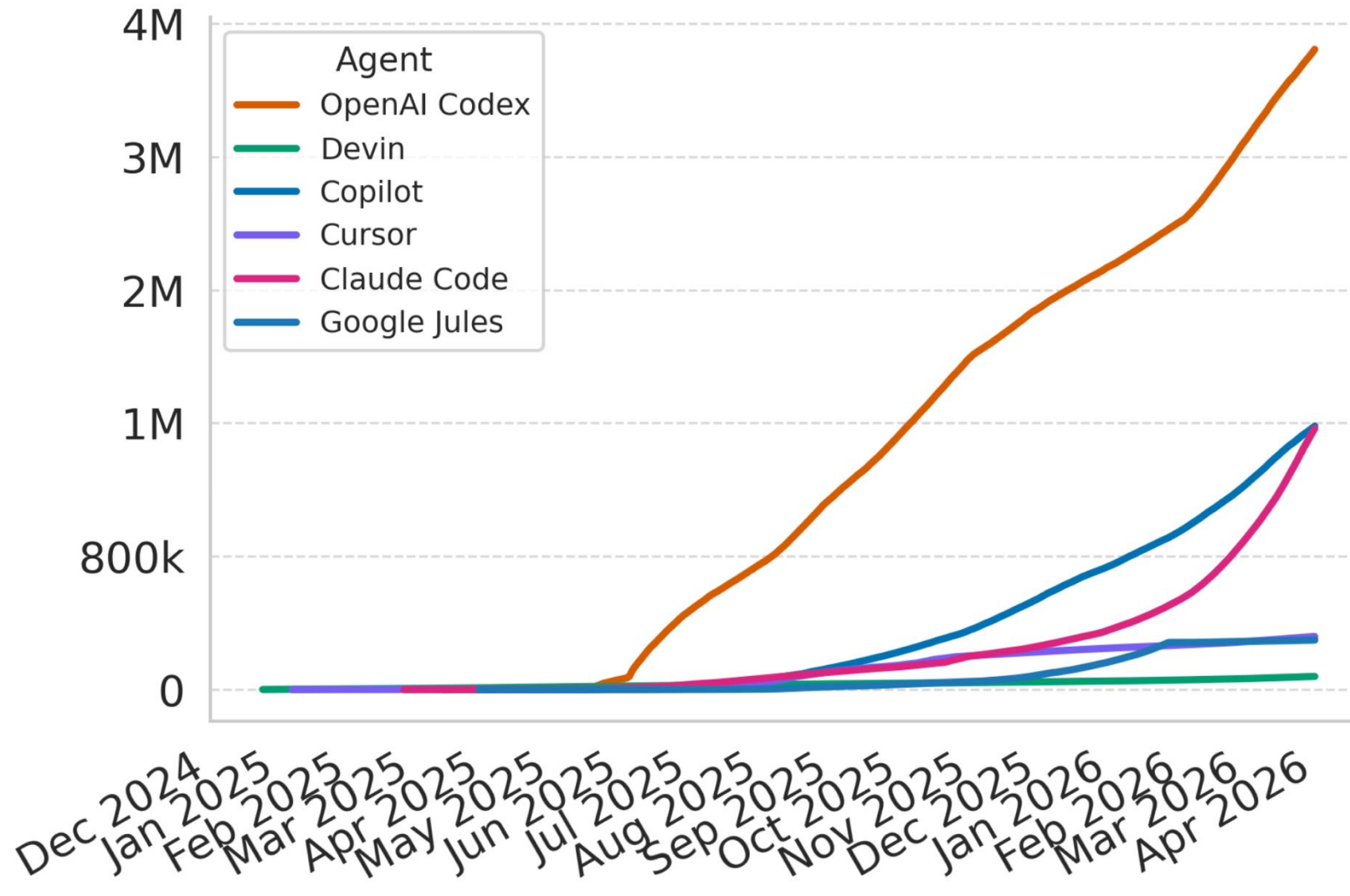
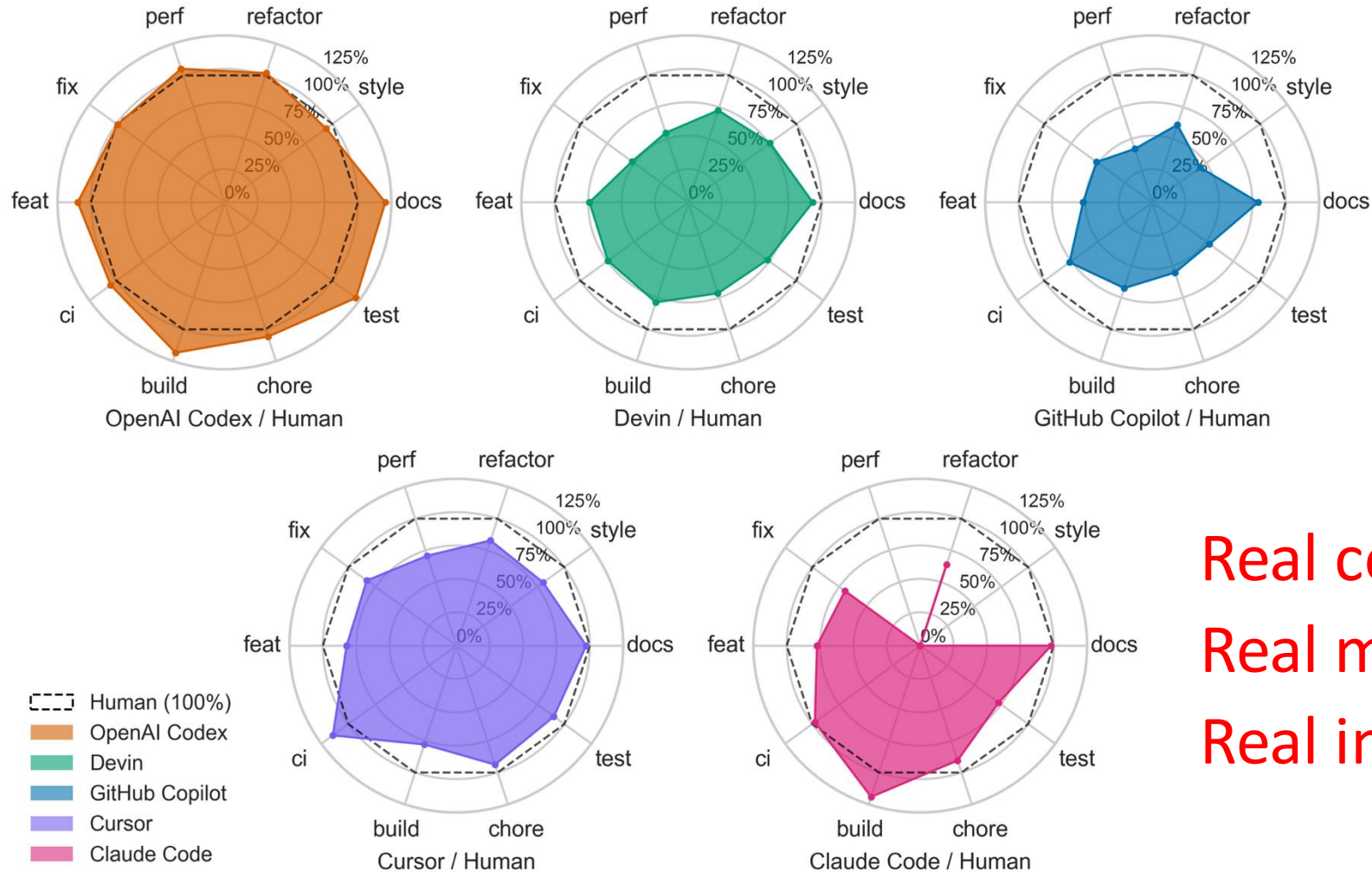


Fig. 1. Structured Agentic Software Engineering (SASE) overview.



Auto. Coding Agents are Not Toys



Real code
Real merges
Real impact

Agentic-PRs acceptance rates match and even exceed human-PRs for various tasks in GitHub projects with 100+ stars

The New Bottleneck: Silence

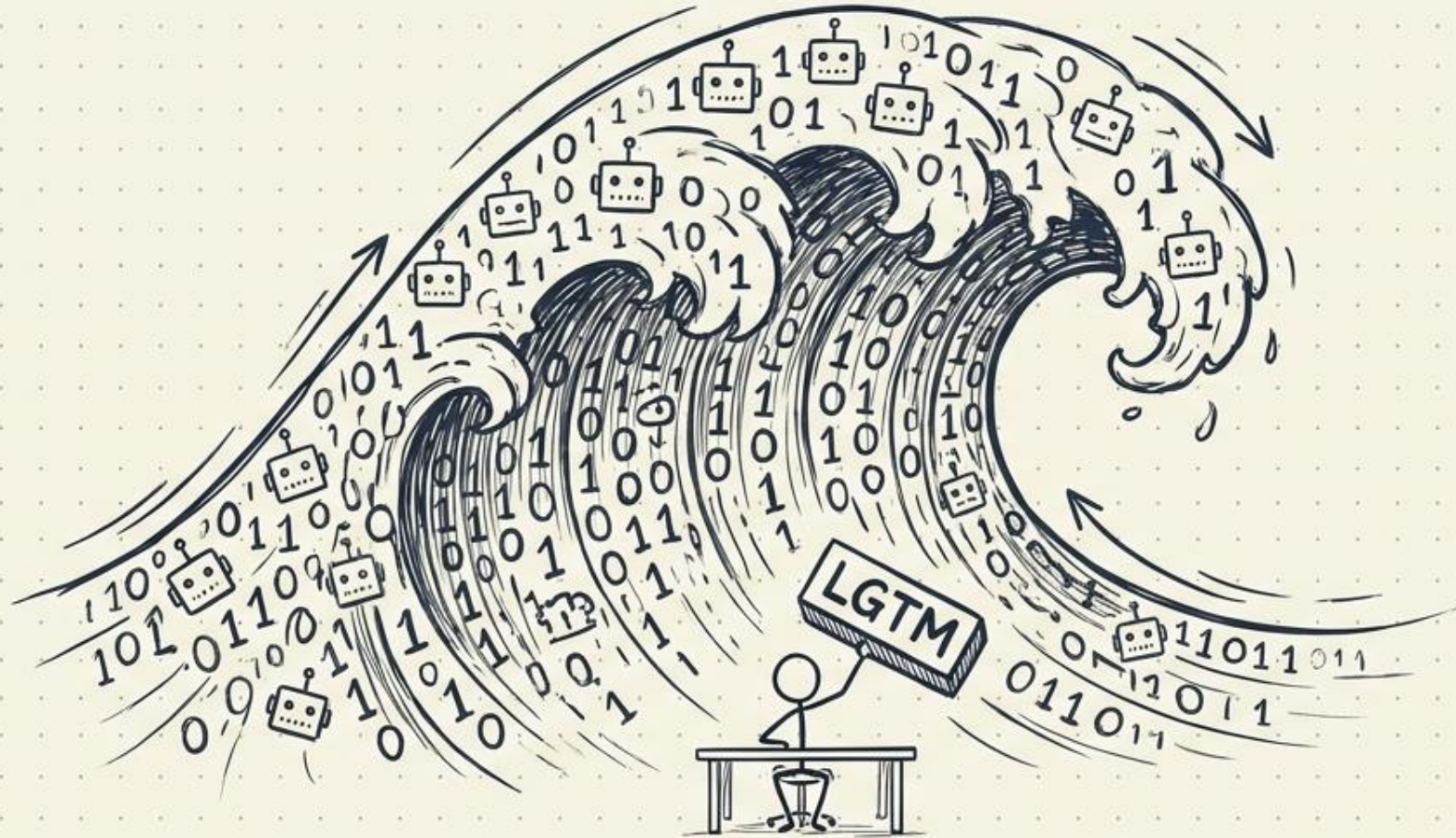


	% All Agents	% Codex	% Copilot	% Cursor	% Devin	% Claude
Unknown (No feedback provided)	34.2	53.8	46.8	38.7	21.0	39.4
Contributor inaction	26.0	5.3	15.3	12.9	39.2	18.2
Submission for verification	16.2	14.0	12.9	9.7	18.7	6.1
Do not add value	9.8	14.5	9.5	12.9	7.5	18.2
Are obsolete	4.1	4.9	5.8	3.2	3.0	6.1
Closed for process reorganization	2.9	1.9	1.1	3.2	3.9	0.0
No confidence in AI-generated code	2.7	2.3	2.2	3.2	3.0	6.1
Non-optimal solutions	1.5	1.2	2.8	6.5	1.1	0.0
Are inactive (author/community)	1.4	0.4	0.9	0.0	2.1	3.0
Lack of project knowledge	0.7	0.7	1.9	6.5	0.2	0.0
High-cost conflict resolution	0.3	0.2	0.6	0.0	0.2	0.0
Are too large	0.2	0.7	0.2	0.0	0.0	3.0
Overly complex implementation	0.1	0.2	0.0	3.2	0.0	0.0
# Total	2418	571	464	31	1319	33

When Agentic-PRs fail, it's not due to bad code, but rather to a lack of feedback and developer inaction

Stop Fooling Yourself: You Cannot Review Your Way Out of the AI Code Flood

The blueprint
for the
Software
Engineering
(SE) Factory



Engineering leaders remain accountable for the risk, failures, and maintenance of code their teams can no longer fully inspect line-by-line.



THE SE FACTORY

HYPER-SPEED WITHOUT LOSING CONTROL

ACCOUNTABILITY

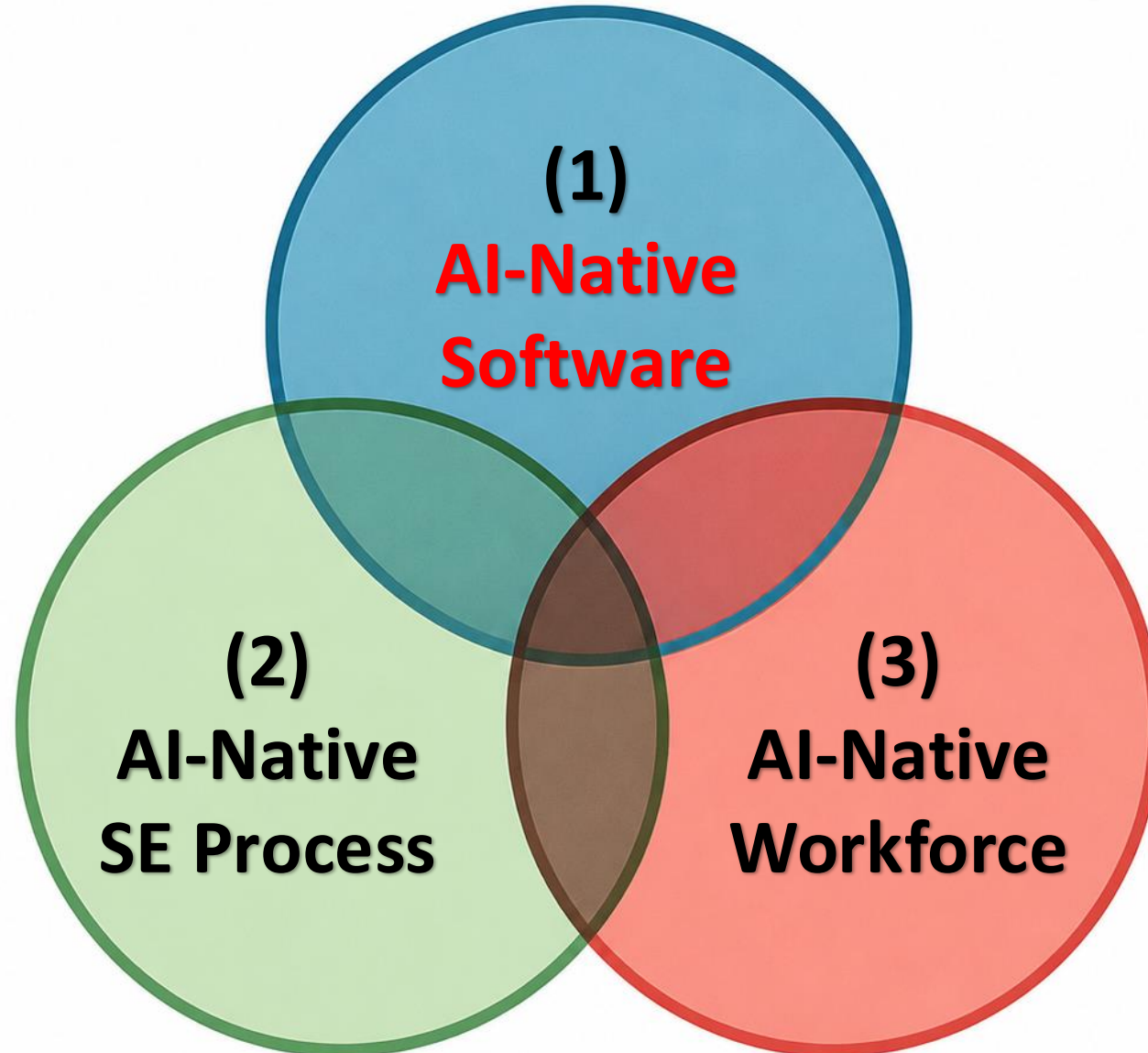
SYMBIOSIS

TRUSTWORTHY
SOFTWARE

COMPOUNDING: EVERY RUN IMPROVES THE NEXT

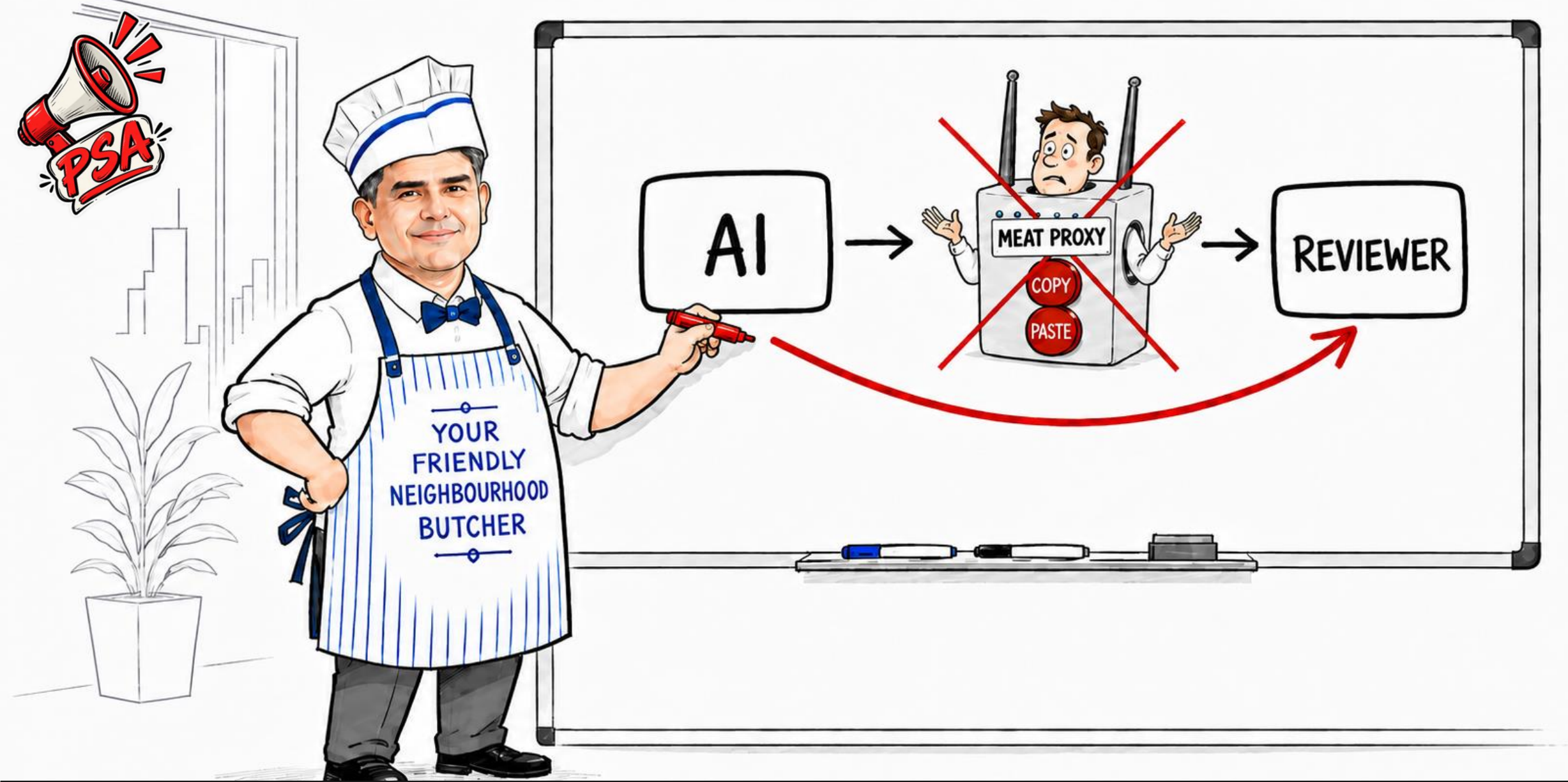


AIware: Software for All by All!



+
(4)
AI-Native
SE Pedagogy

<https://www.aiwareconf.org/>
<https://www.aiwarebootcamp.io/>



**USE AI AGGRESSIVELY; REMAIN THE ACCOUNTABLE OWNER
IF YOUR VALUE IS FORWARDING AI OUTPUT, THE WORKFLOW WILL ROUTE AROUND YOU.**

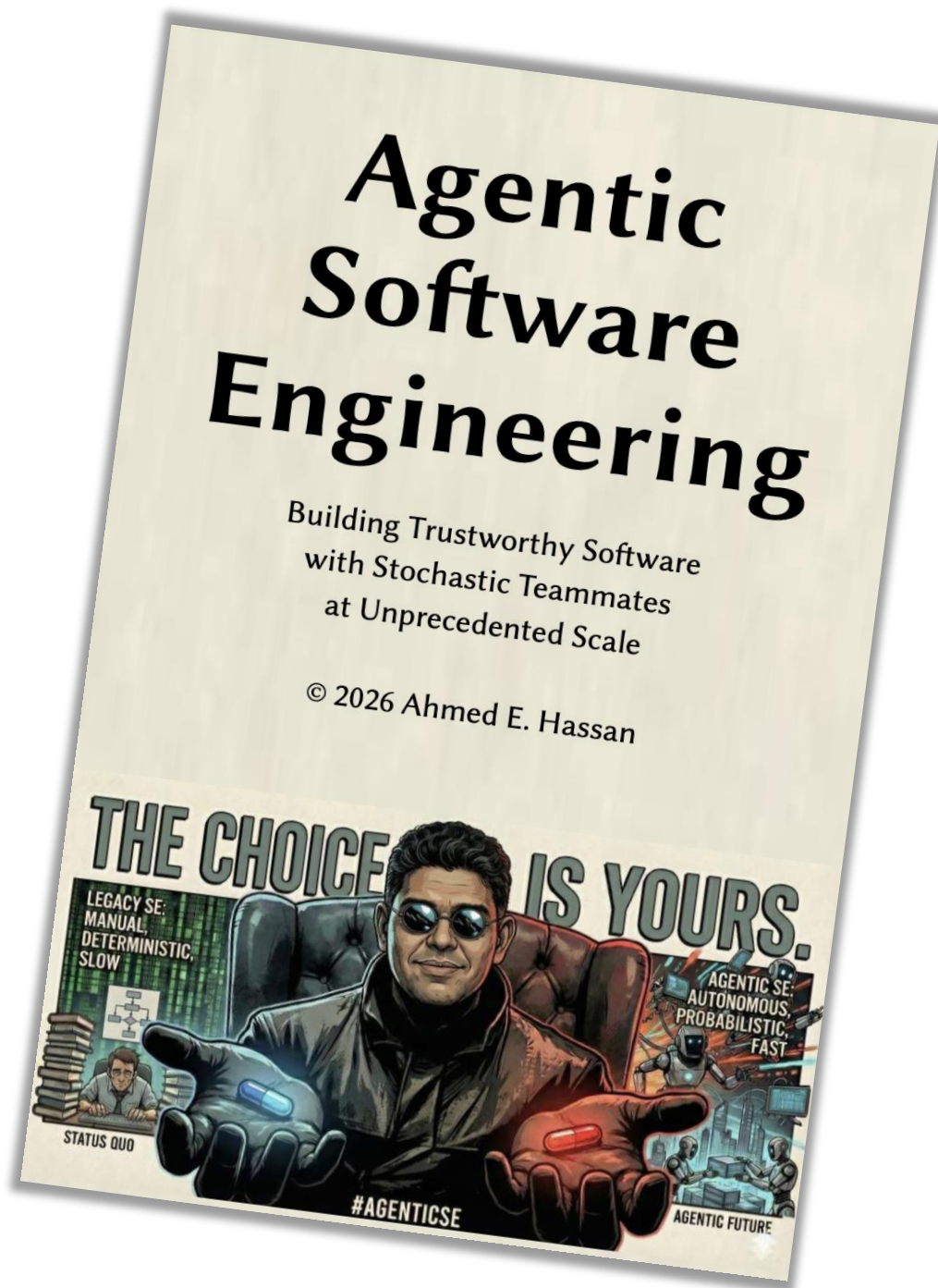


***A fool with a tool
is still a fool!***

<https://agenticse-book.github.io/>



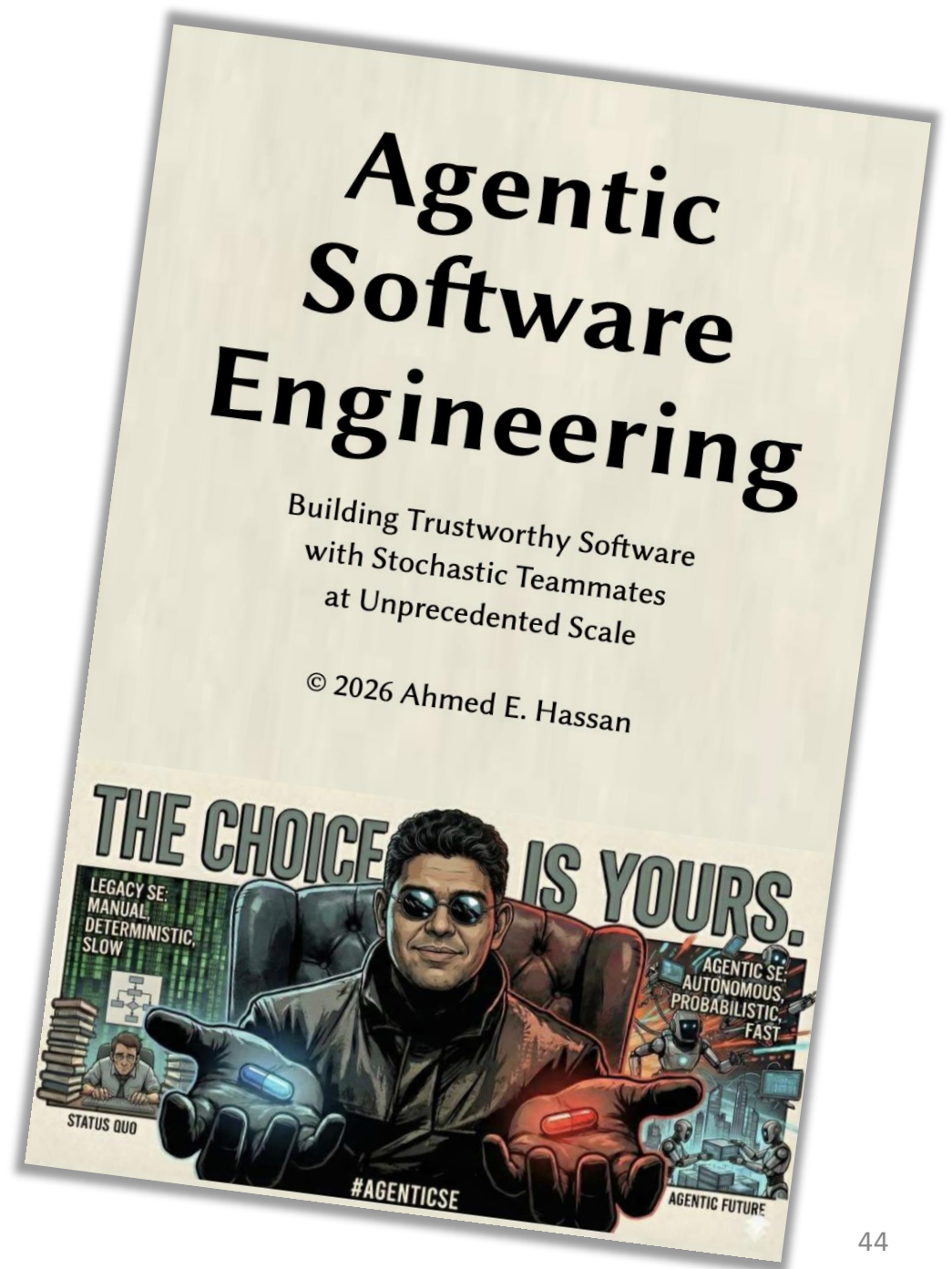
Follow the white rabbit!



“ Engineering has never been the art of assembling perfect parts. Civil engineers do not start with flawless steel, flawless concrete, and flawless workers. Materials have tolerances. Welds can fail. Teams make mistakes.

What makes a bridge safe is not magical steel. It is the system: redundancy, safety margins, inspection regimes, standardized practices, and a relentless focus on failure modes. ”

<https://agenticse-book.github.io/>





*I GUESS YOU GUYS AREN'T
READY FOR THAT YET.
**BUT YOUR KIDS ARE
GONNA LOVE IT.***

— MARTY McFLY —