

SoftEd GenAI Day · May 14, 2026

Stop Using a Sledgehammer

Matching AI autonomy to the task at hand



Jess Wolfe · Customer Success, Swarmia

***Everyone's adopted the tools.
Almost no one has adopted a theory.***

60%

of work has AI integrated into it

80–100%

of tasks still need human oversight



Source: Anthropic Agentic Coding Trends Report

Two ways to misuse AI



Sledgehammer at a thumbtack

Firing up an autonomous agent to fix a typo.

High autonomy, low-stakes task.



Wrench on a screw

Tab-completing through a refactor an agent could have shipped overnight.

Low autonomy, high-leverage task.



JESS WOLFE

Customer Success · Swarmia

Hi, I'm Jess 🖐️

Customer Success Manager at Swarmia.

WHAT IS SWARMIA?

Engineering intelligence platform.

Engineering teams use Swarmia to connect GitHub, Jira, GitLab and AI tooling in one place — so leaders see who's using which AI tool, on which projects, with which results.

AI impact

track adoption & ROI

Investment balance

& capex / opex automation

DORA + SPACE

engineering performance

The five levels of AI autonomy



Miikka Holkeri, Swarmia — March 2026

L1		Assistive <i>GitHub Copilot · Cursor Tab · Grammarly · Gmail Smart Compose</i>	Autocomplete. You drive.
L2		Conversational <i>ChatGPT · Claude · Microsoft Copilot · Google Gemini</i>	Multi-context chat. You steer; model proposes.
L3		Task agent <i>Claude Code · ChatGPT Agent · Microsoft Copilot agents · Salesforce Agentforce · Cursor background agents</i>	Hand off a defined task. Get back a reviewable result.
L4		Autonomous teammate <i>Dependabot · Salesforce Einstein · Claude CoWork · Microsoft Copilot scheduled agents · Zapier AI triggers</i>	Work shows up without you asking.
L5		Agentic avalanche <i>Claude sub-agents · AutoGen · CrewAI · LangGraph · custom workflows</i>	Many agents in parallel.

INDUSTRY INSIGHT

<2% of mid-to-large orgs operate above L3. Most tools span L1-L3 — the level is set by how you use them.

Higher is not always better. · Clarity of done is your Definition of Ready for AI.

Picking a level is a staffing decision.

Domain expert on intern work

→ *wasted judgment.*

Middle-schooler running the company

→ *disaster.*

The right person on the right task

→ *leverage.*






AI is the same. The question isn't how autonomous can this be — it's how autonomous should it be?

EXAMPLE A

The vague spec

THE TASK

Improve the signup flow conversion rate.

- L1  Assistive
- L2  Conversational
- L3  Task agent
- L4  Autonomous teammate
- L5  Agentic avalanche



What level is this? Drop your guess in chat — L1 / L2 / L3 / L4 / L5

The vague spec

L2

CONVERSATIONAL

THE TASK

Improve the signup flow conversion rate.

THE RULE

If the spec fits on a sticky note and could mean three different things, you are at Level 2.



INDUSTRY INSIGHT






~70% of these are still being worked at L1 — engineers tab-completing through ambiguous specs.

EXAMPLE B

Release notes

THE TASK

You need customer-facing release notes for the latest release. Group by feature area, classify each as new / improved / fixed, skip anything tagged internal-only, use our standard format. Output: ready for editorial pass.

- L1  **Assistive**
- L2  **Conversational**
- L3  **Task agent**
- L4  **Autonomous teammate**
- L5  **Agentic avalanche**



What level is this? Drop your guess in chat — L1 / L2 / L3 / L4 / L5

Release notes

L3

TASK AGENT

THE TASK

You need customer-facing release notes for the latest release. Group by feature area, classify each as new / improved / fixed, skip anything tagged internal-only, use our standard format. Output: ready for editorial pass.

THE RULE

If you can describe done in one sentence and a reviewer can verify it in a single pass - promote it.



INDUSTRY INSIGHT






This is the level most teams under-use. The L3 adoption gap - work that should be L3 but is not - is where most AI leverage hides.

EXAMPLE C

Support ticket triage

THE TASK

A customer submits a support ticket. The agent reads the ticket, checks account tier and recent activity, classifies urgency, drafts a first response from the knowledge base, and escalates anything high-risk or low-confidence. No one manually starts the workflow.

- L1  **Assistive**
- L2  **Conversational**
- L3  **Task agent**
- L4  **Autonomous teammate**
- L5  **Agentic avalanche**



What level is this? Drop your guess in chat — L1 / L2 / L3 / L4 / L5

Support ticket triage

L4

AUTONOMOUS

THE TASK

A customer submits a support ticket. The agent reads the ticket, checks account tier and recent activity, classifies urgency, drafts a first response from the knowledge base, and escalates anything high-risk or low-confidence. No one manually starts the workflow.

THE RULE

If the same trigger keeps happening and the acceptable response is stable, make the agent the initiator with clear escalation rules for exceptions.



INDUSTRY INSIGHT






Less than 2% of mid-to-large orgs operate here today. Where the leverage gap lives.

EXAMPLE D

The PM with a hunch

THE TASK

What if the dashboard opened with a plain-English weekly summary: what changed, what looks promising, and what question a manager should ask next?

- L1  Assistive
- L2  Conversational
- L3  Task agent
- L4  Autonomous teammate
- L5  Agentic avalanche



What level is this? Drop your guess in chat — L1 / L2 / L3 / L4 / L5

The PM with a hunch

THE TASK

What if the dashboard opened with a plain-English weekly summary: what changed, what looks promising, and what question a manager should ask next?

THE RULE

If you are still discovering what to build, stay at L2. The output is learning — often used for a sharper Level 3 task later.



INDUSTRY INSIGHT






PM-driven prototyping is the fastest-growing AI use case — and one of the most under-supported by leadership.

EXAMPLE E

Software capitalization at scale

THE TASK

4,000 pull requests and 6,000 Jira issues last quarter. Finance needs all work classified as capitalizable or operating expense. Today: a month of report-building, meetings, defensive estimates, audit pushback.

- L1  Assistive
- L2  Conversational
- L3  Task agent
- L4  Autonomous teammate
- L5  Agentic avalanche



What level is this? Drop your guess in chat — L1 / L2 / L3 / L4 / L5

Software capitalization at scale

THE TASK

4,000 pull requests and 6,000 Jira issues last quarter. Finance needs all work classified as capitalizable or operating expense. Today: a month of report-building, meetings, defensive estimates, audit pushback.

THE RULE

L5 is not about courage — it is about control. If you cannot define the policy, evidence, confidence threshold, exception path, and audit trail, you are not ready for L5.



INDUSTRY INSIGHT

Most companies still do this in spreadsheets and meetings. The opportunity is defensible classification, cleaner evidence, and fewer audit surprises.

The bottleneck has moved.

Quality of intent

is the new bottleneck.

Clarity of done is your Definition of Ready for AI.

AI is an amplifier. Adopt without clarity and you scale your problems, not your output.

Before you reach for an AI tool, ask...

FIRST

Can I describe what done looks like?

If no — this isn't an AI problem. It's a clarity problem.

THEN ASK



Can I describe done in one sentence with no ambiguity?

Level 3+



Will the same trigger keep happening?

Level 4



Is volume too high for manual review — and do I have a quality control?

Level 5



If none of the above — I am at Level 2, and that's fine.

Level 2

Stop swinging a sledgehammer. Match the tool to the task.

And invest in the quality of your intent — that's where the leverage lives now.

Want to see your team's AI adoption in Swarmia?

AI adoption is one of several things Swarmia helps engineering leaders see.



AI tool adoption

Who's using which AI tool, where it's actually helping.



Engineering effectiveness

Cycle time, review time, throughput, with team-level cuts.



Investment & ROI

Where engineering time goes — features, growth, paying down debt.



Team health & flow

Sustainable pace, on-call burden, retention signals.



**Book a 30-minute meeting with
Otto Hilska, CEO Swarmia**

See your team's AI adoption in Swarmia. No commitment.

Questions?

Jess Wolfe

jess@swarmia.com

LinkedIn:

<https://www.linkedin.com/in/thejessicawolfe/>



The framework:

swarmia.com/blog/five-levels-ai-agent-autonomy

