

Where AI Helps and Where Humans Still Lead in L2 and L3 IT Support

Raghunath Reddy Koilakonda, Independent Researcher, Texas, USA, raghunath.koilakonda@gmail.com

Abstract—AI tools are starting to move from help-desk chat into the harder work of Level 2 (L2) and Level 3 (L3) IT support. This paper describes where AI can save time in these tiers, such as routing cases, summarizing history and suggesting fixes, and where human judgment is still needed. It proposes a simple human-in-the-loop workflow, outlines the main risks, and suggests starting with a small, controlled pilot.

Keywords—IT service management, technical support, artificial intelligence, large language models, incident management, human-in-the-loop

I. INTRODUCTION

IT support is usually organized in tiers. Level 1 (L1) handles common requests such as password resets and basic how-to questions. L2 takes cases that need deeper troubleshooting, such as configuration faults, integration errors and recurring incidents. L3 involves engineers or product specialists who look for root causes, fix defects and change the system itself.

Researchers have begun testing AI on this upper-tier work. A study at Microsoft evaluated large language models on more than 40,000 cloud service incidents from over 1,000 services, and included feedback from the engineers who owned those incidents [1]. For support leaders, the question is less whether to use AI and more where it helps and where it does not.

II. WHERE AI SPEEDS THINGS UP

Much of the time in L2 and L3 goes to understanding the problem, not solving it. AI can help most with that reading, sorting and drafting work.

- **Sorting and routing.** AI can read a new case, guess the product area and urgency, and send it to the right team.
- **Summarizing case history.** A long thread of notes, emails and attachments can be reduced to a short summary of what has already been tried.
- **Suggesting fixes.** By comparing a new case with past ones, AI can propose likely causes and next steps. In the Microsoft study, fine-tuning language models on historical incident data made them more effective at this task [1].
- **Spotting duplicates.** Several cases reporting the same fault can be grouped into one incident, which helps a team see a problem earlier.
- **Drafting communication.** First drafts of customer updates, hand-over notes and post-incident summaries save writing time, as long as a person reviews them before they go out.
- **Finding relevant documentation.** Assistants that search a knowledge base are also being studied for incident handling. One 2024 short paper evaluates such an assistant using escalation rate, first-contact resolution and average resolution time, though its results are preliminary [2].

These uses share a pattern: AI prepares information and options, and a person makes the decision. They are also the lowest-risk place to begin.

In my own L2 work, the clearest time-saver would be incident summarization and root-cause tracking. A long case can run through many updates, handovers and attachments, and a short summary of what has been tried and what the likely cause is saves the next analyst from reading it all. AI could also help when a new incident arrives, by showing similar past incidents and the fixes that worked, so the analyst has a starting point instead of a blank page. I would not rely on AI for the conclusion, though. It can produce answers that sound right but are wrong or incomplete for the specific incident, so I treat its output as a draft to check, not a diagnosis. For that reason, it needs guardrails: it should show which past incidents it drew on, and an engineer should confirm the root cause and approve any fix before it is applied. With those controls in place, the benefits in time saved are still substantial.

III. WHERE HUMANS ARE STILL NEEDED

AI is less reliable where cases are unusual, where context is missing, or where a wrong answer is costly. Those conditions are common in L2 and L3.

- **New or multi-system faults.** When a problem spans several systems or has no precedent in past cases, an AI suggestion rests on weak evidence and can still sound confident.
- **Judgment about customer impact.** Deciding whether to escalate, offer a workaround or accept a missed service target involves business context and relationships that AI may not see.
- **Security- and data-sensitive actions.** Changing access, restoring data or touching production systems should need human approval and a clear record of who approved it.
- **Root-cause analysis.** AI can offer hypotheses, but engineers have to test them, confirm the cause and decide on a lasting fix.
- **Checking AI output.** Every suggestion needs a reviewer who knows the system well enough to tell a good answer from a plausible one. AI does not remove the need for skilled staff.
- **Dealing with upset customers.** Empathy, negotiation and trust-building remain human work.

Models built from past incidents can also go stale as systems change. The Microsoft authors note that such models would need frequent retraining on recent incident data [3].

IV. A PRACTICAL HUMAN-IN-THE-LOOP WORKFLOW

The safest design keeps a qualified person in control of every action that changes a customer's service or a production system. A simple workflow looks like this:

1. **Intake.** AI classifies the new case, suggests a priority and routes it to the right team.
2. **Preparation.** AI summarizes the history, lists similar past cases and suggests possible causes, each with a link to its source.
3. **Review.** The engineer reads the suggestions, checks them against the evidence and decides what to do. A suggestion with no source is treated as unverified.
4. **Action.** The engineer carries out or approves the fix. AI may draft the customer update, which the engineer edits and sends.
5. **Feedback.** The engineer marks whether the suggestion helped. Confirmed fixes go into the knowledge base for future cases.

The feedback step is what lets the system improve. Without it, the same weak suggestions keep coming back.

V. RISKS AND GOVERNANCE

- **Wrong but convincing answers.** Ask for sources with every suggestion and train staff to challenge them.
- **Data privacy.** Case records often contain customer or personal data. Decide what may be sent to an AI service and check how the provider stores and uses it.
- **Over-reliance.** If engineers stop practicing diagnosis, their skills can fade. Keep some cases human-led and review samples of AI-assisted work.
- **Weak audit trail.** Record what AI suggested, what the engineer decided and why, especially for security or customer-impacting actions.
- **Outdated knowledge.** Review and retire old knowledge articles so AI does not repeat old mistakes.

The NIST AI Risk Management Framework is a voluntary guide for managing AI risk. It organizes the work into four functions: govern, map, measure and manage [4]. A support team can borrow that structure even for a small pilot, starting with a named owner, a short written policy and regular quality reviews.

VI. HOW TO START

Begin with a small, low-risk pilot.

1. Pick one or two tasks, such as case summaries and routing suggestions.
2. Record a baseline first: average time to resolve, how often cases are reassigned, and engineer time spent reading case history.
3. Run the pilot with a small group and compare against the baseline.
4. Collect engineer feedback on what was useful and what was wrong.
5. Train staff on how to review AI output and when not to use it.
6. Expand only the tasks that show clear value.

VII. CONCLUSION

AI is a useful assistant in L2 and L3 support when it prepares information, drafts text and suggests options. It is a poor substitute for engineering judgment, customer empathy and accountability for decisions. The best results come from a clear split of work: AI handles reading, sorting and drafting, and people handle diagnosis, approval and relationships. Teams that start small, measure results and keep humans in control can gain speed without losing quality or trust.

REFERENCES

- [1] T. Ahmed, S. Ghosh, C. Bansal, T. Zimmermann, X. Zhang, and S. Rajmohan, "Recommending root-cause and mitigation steps for cloud incidents using large language models," in *Proc. Int. Conf. Softw. Eng. (ICSE)*, 2023, arXiv:2301.03797. [Online]. Available: <https://arxiv.org/abs/2301.03797>.
- [2] R. Schmidt, R. Alt, and A. Zimmermann, "Generative assistants for handling incidents in IT service management," in *Proc. Int. Conf. Inf. Syst. (ICIS)*, 2024, Paper 19. [Online]. Available: https://aisel.aisnet.org/icis2024/digtech_fow/digtech_fow/19.

[3] "Large-language models for automatic cloud incident management," Microsoft Research Blog. [Online]. Available: <https://www.microsoft.com/en-us/research/blog/large-language-models-for-automatic-cloud-incident-management>.

[4] E. Tabassi, "Artificial Intelligence Risk Management Framework (AI RMF 1.0)," Nat. Inst. Standards Technol., Gaithersburg, MD, USA, Rep. NIST AI 100-1, 2023, doi: 10.6028/NIST.AI.100-1.