

EXPECTED ESCAPED HARM, TERM BY TERM

# Reliability-Equation Worksheet

Reliability is not accuracy. Estimate each of the four terms for one real agent action, give a range and a confidence, name an owner, then read the product as expected escaped harm.

**BRING** One agent action you are deciding whether to trust \_\_\_\_\_

## The four terms

Estimate each term, give the range you would defend, mark how confident you are, and name who owns getting the number right. Writing "unknown" is a valid and valuable answer: it shows where the architecture has been set to one without anyone deciding.

Term	Plain-language question	Estimate	Range	Confidence	Owner
P(error): how often does the system make a materially relevant mistake on the real task?					
P(undetected   error): what structurally different check catches the error before it becomes an action?					
P(uncontained   error, undetected): what hard boundary stops an uncaught error reaching the live system?					
E(severity   escaped): in what units, and how large, is the harm once it lands?					

---

## Notes - read the product, not the slide

Multiply the four terms for expected escaped harm per unit of work. The terms are not independent, so treat the result as an ignorance map, not a verdict. Write which factor your roadmap is improving, and which control would move which term.

---

---

---

---

---

---

---

---

---

---

Companion worksheet to **Essay 05 · The Reliability Equation**, in the series **Architecting the AI Coworker**. · Dr Peter McCann Strain  
· Fill this in against one real agent, action class or vendor. © 2026 Peter McCann Strain.

Series Companion + all 22 worksheets: **Release\_v12/Series\_Companion.pdf**