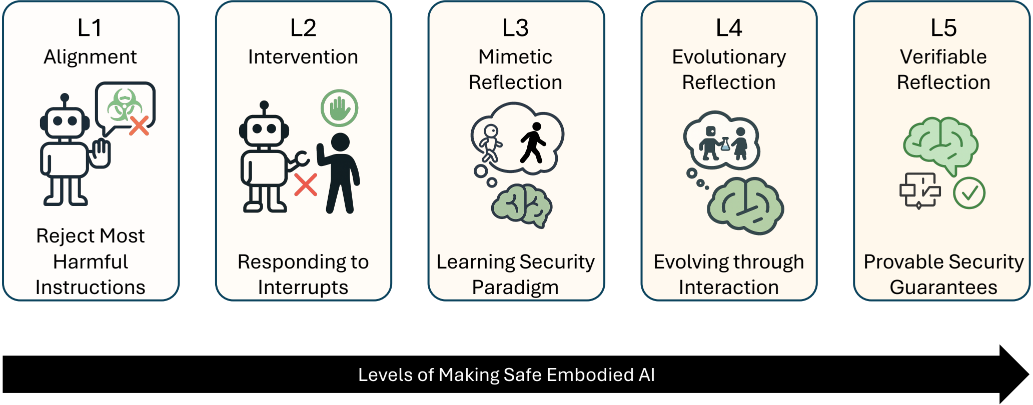

L1
Alignment
Reject Most Harmful Instructions

L2
Intervention
Responding to Interrupts

L3
Mimetic Reflection
Learning Security Paradigm

L4
Evolutionary Reflection
Evolving through Interaction

L5
Verifiable Reflection
Provable Security Guarantees

Levels of Making Safe Embodied AI