# Code Nobody Wrote: Running Agents You Can’t Trust

*Todd Neal, AWS*

# Reflections on Trusting Trust



> The moral is obvious. You can't trust code that you did not totally create yourself. … No amount of source-level verification or scrutiny will protect you from using untrusted code.
>
> — Ken Thompson (1983)

# Secure Computing Systems



> The only truly secure system is one that is powered off, cast in a block of concrete and sealed in a lead-lined room with armed guards -- and even then I have my doubts.
>
> — Eugene H. Spafford (1989)

# Containment



## Kernel

- Virtual Machine
- Hardened (SELinux)

## Resource Limits

- CPU / memory
- Disk / Network I/O
- Conntrack tables

# Actions



## Custom Shells

- Strands Box
- OpenShell
- nono

## Policy

- Cedar/Dogwood
- Rego

## MCP Servers

- Restricted Actions

# Reach



## Network

- None - KEP-6313 – Non-Network Pods
- Proxied

## Credentials

- None
- MCP Servers
- Injected via Proxy

# Where Are We Going?



It is difficult to make predictions, especially about the future.

— Origin Uncertain