Responsible Research and Innovation in System Design

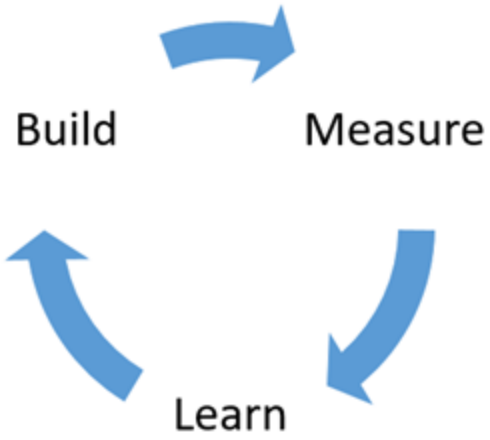
Dependable and Deployable AI for Robotics 2

Barbara Webb

- The Collingridge dilemma: *"attempting to control a technology is difficult…because during its early stages, when it can be controlled, not enough can be known about its harmful social consequences to warrant controlling its development; but by the time these consequences are apparent, control has become costly and slow"* (Collingridge, 1980: 19).
- Can we incorporate actions in our design process to try to minimise harmful social consequences?

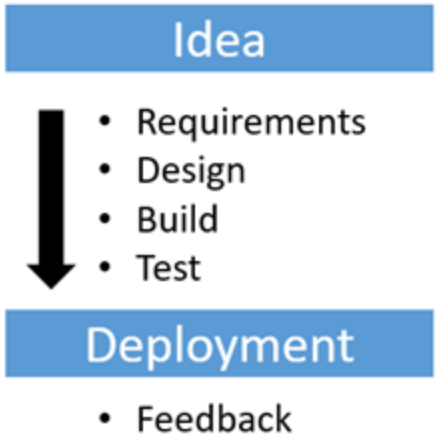
There are many alternative design approaches:

Agile

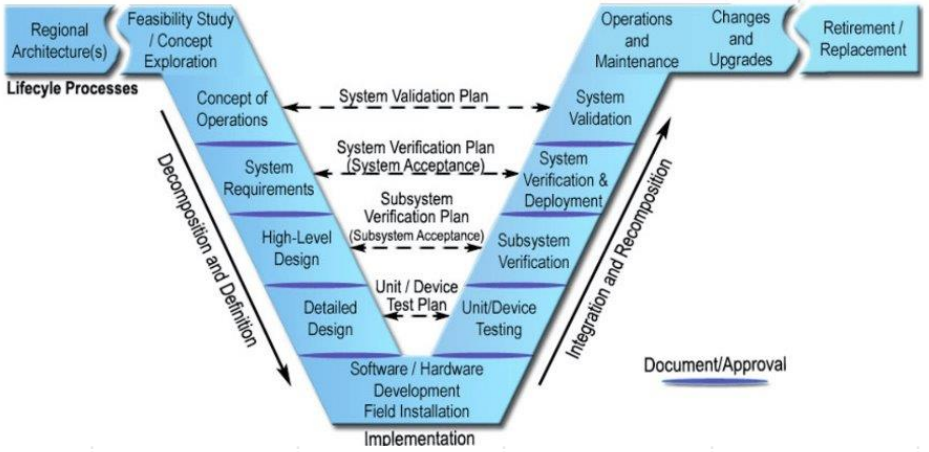


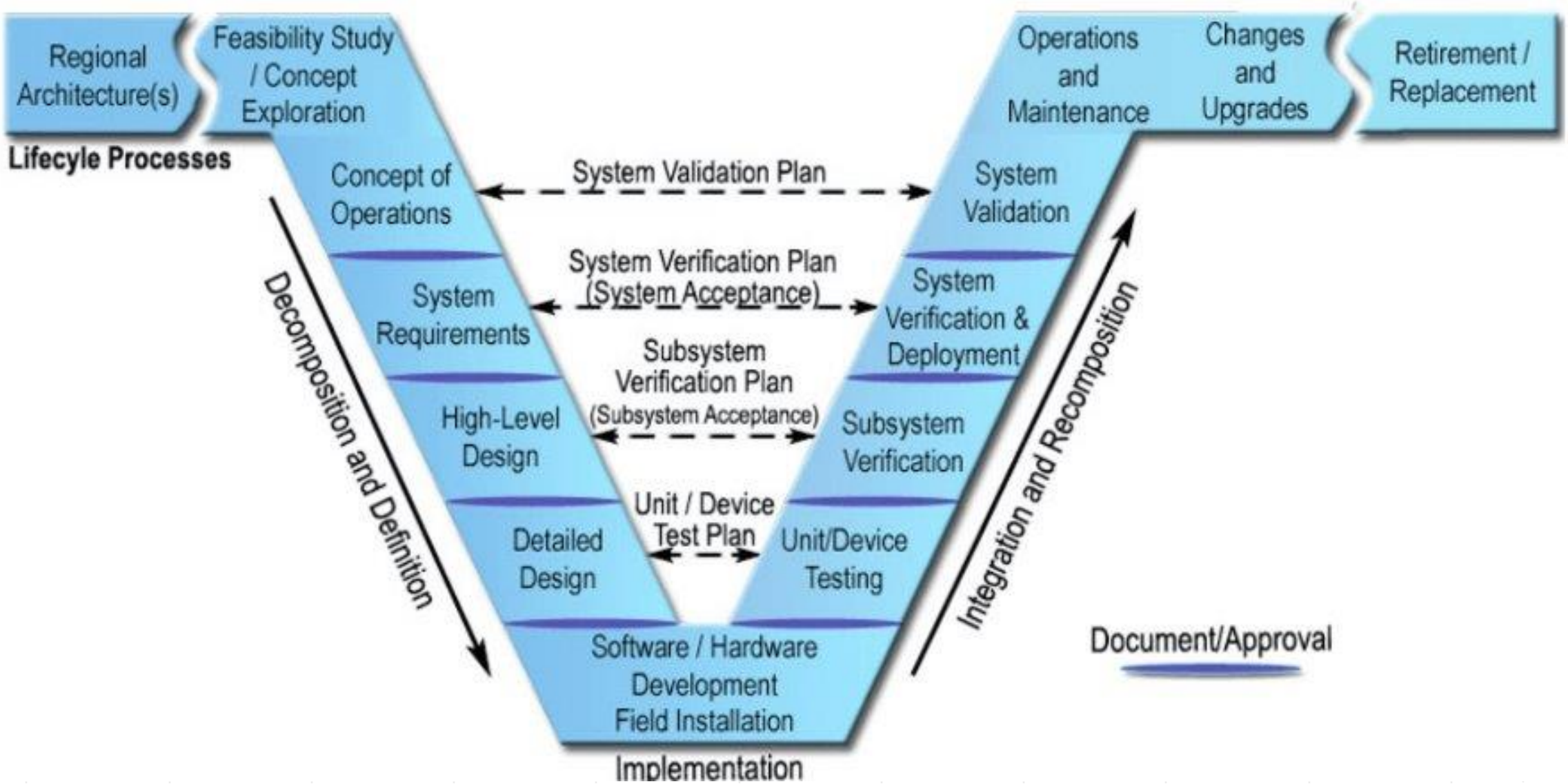
VS

Waterfall

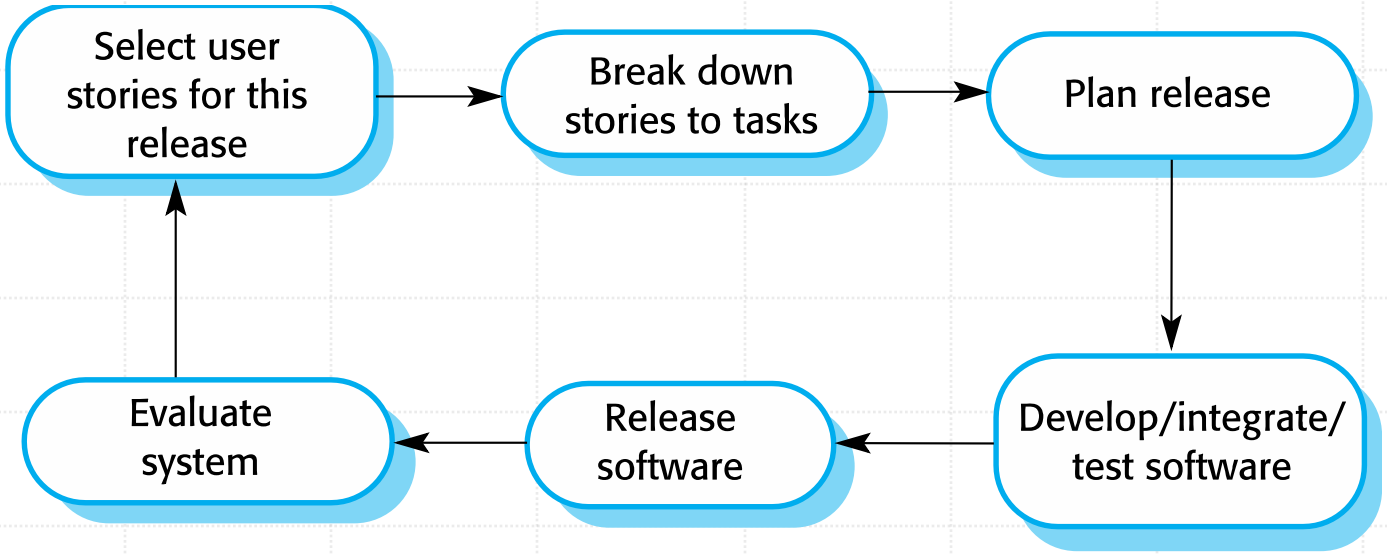


V-model




Regional Architecture(s)
Feasibility Study / Concept Exploration
Operations and Maintenance
Changes and Upgrades
Retirement / Replacement
Lifecyle Processes
Concept of Operations
System Validation Plan
System Validation
System Requirements
System Verification Plan (System Acceptance)
System Verification & Deployment
Subsystem Verification Plan (Subsystem Acceptance)
High-Level Design
Subsystem Verification
Unit / Device Test Plan
Detailed Design
Unit/Device Testing
Decomposition and Definition
Integration and Recomposition
Software / Hardware Development Field Installation
Implementation
Document/Approval

The extreme/agile release cycle



Some principles of agile:

- Transparency: team responsibility to share information and team member responsibility to stay informed
- Inspection: both of what is produced, and of the process itself, to iteratively improve
- Adaptability: to the reality you are actually dealing with, rather than imagined scenario

User Stories

- Instead of Use Cases, Agile project owners do "user stories"
  - **Who** (user role) – Is this a customer, employee, admin, etc.?
  - **What** (goal) – What functionality must be achieved/developed?
  - **Why** (reason) – Why does user want to accomplish this goal?

As a [user role], I want to [goal], so I can [reason].

- Example:
  - “As robot A, I want to be able to kick the ball towards robot B, so that B can score a goal"

- Collection of stories for the whole project is the ‘product backlog’

Human-centred design

See also user-centred, participatory design, value-sensitive design

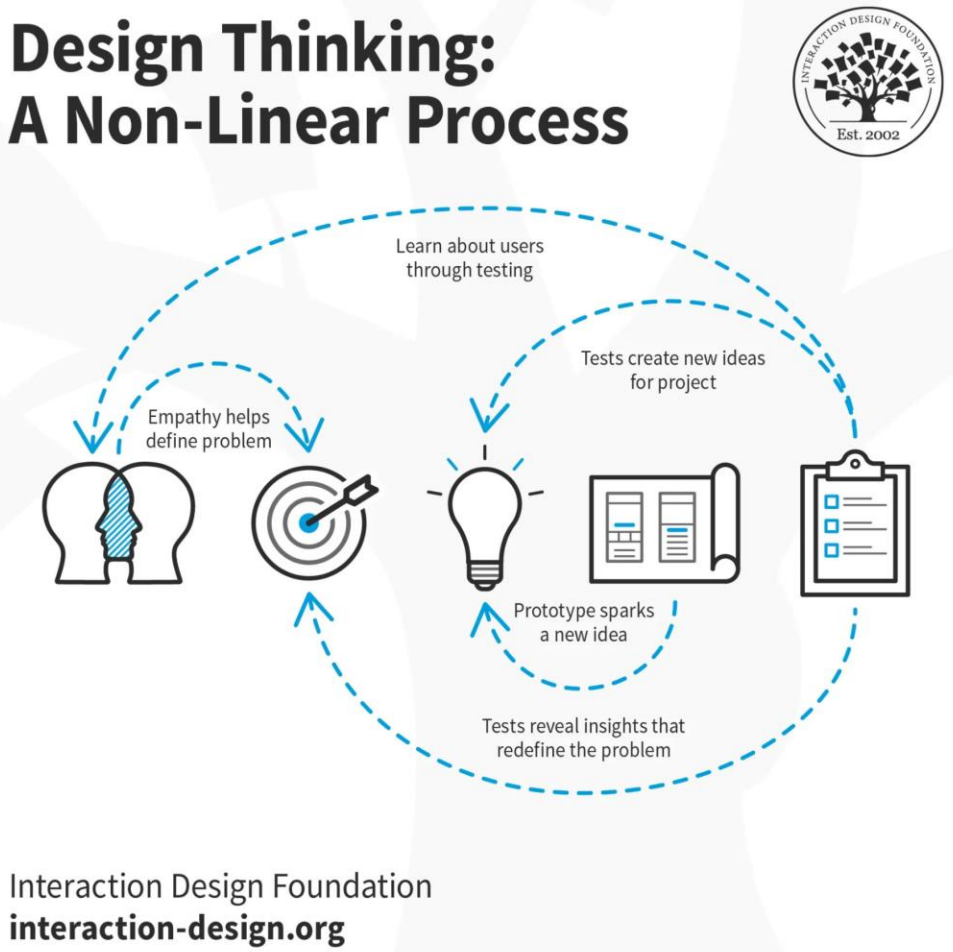
Four key principles:

- **People-centered**: Focus on people and their context to create appropriate solutions.
- **Understand and solve the right problems:** Understand the underlying fundamental issues.
- **Everything is a system:** Consider everything as a system of interconnected parts.
- **Small and simple interventions:** Do iterative work and get constant feedback.

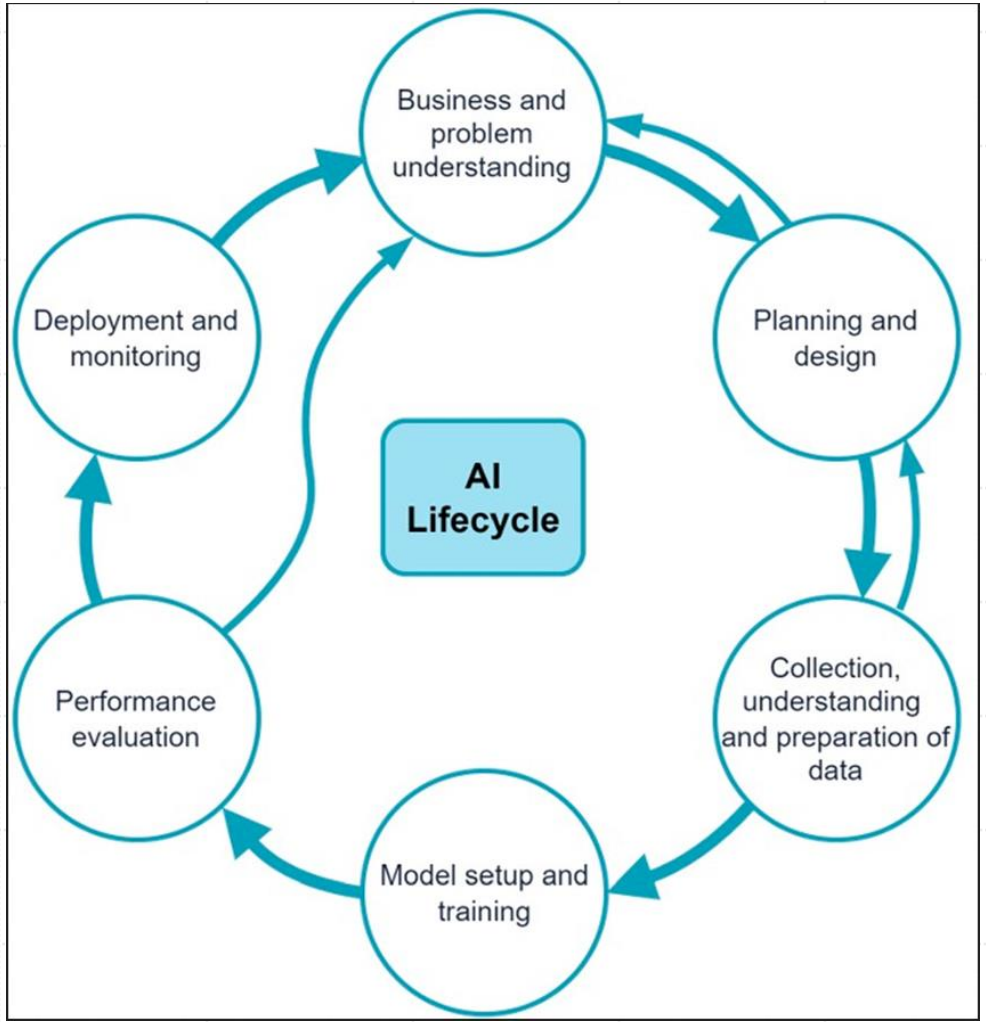
Human-centred design

Five steps:

- **Empathize**: understand the user's point of view in depth; avoid assuming you know what the problem is.
- **Define**: analyze the core issues from step one to produce a clear problem statement..
- **Ideate**: Don't fixate on first solution but use brainstorming and other creativity tools to expand the solution space..
- **Prototype**: Create scaled down versions to test before sinking resource into full-scale implementation.
- **Test**: re-engage with the user



RRI in the AI (ML) lifecycle



- Ortega-Bolaños, R., Bernal-Salcedo, J., Germán Ortiz, M. *et al.* Applying the ethics of AI: a systematic review of tools for developing and assessing AI-based systems. *Artif Intell Rev* **57**, 110 (2024). https://doi.org/10.1007/s10462-024-10740-3

Critical design

- Creatively explore what might be
- Expressed as fictions, objects to think with, etc.
- Point is not useful or efficient engineering but to create a critical social context for the intended system
- For example, take goal of a project to its extreme, revealing how solving one problem might create a variety of others
- But sometimes reality is stranger than fiction…

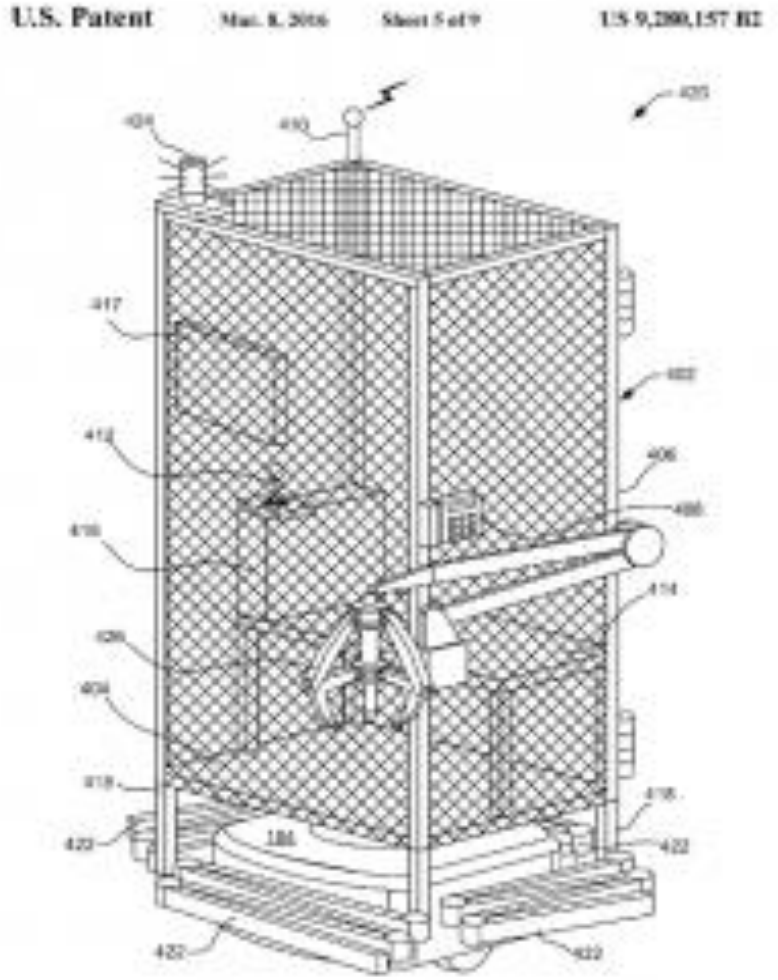


Illustration from a 2013 Amazon patent, “System and method for transporting personnel within an active workspace.”