Exploring the Unintended Consequences of Automation in Software

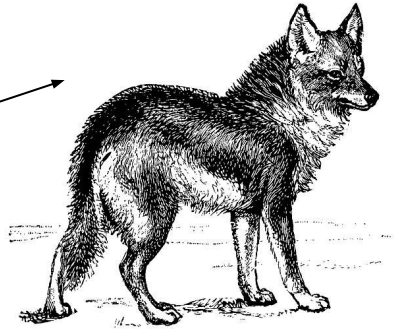
Courtney Nash
The VOID

SRECon EMEA
Oct 29, 2024

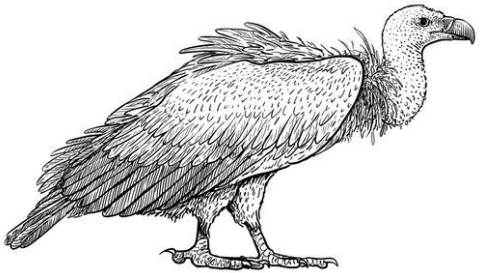
An estimated 500,000 people in India died over the course of 5 years due to the sudden loss of their vulture population.

Other predator populations increase

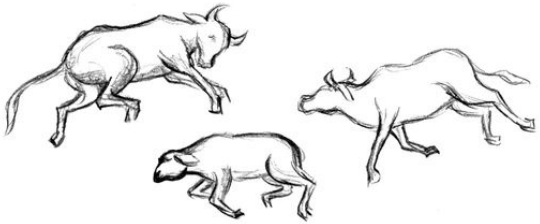
Vultures disappear



Disease rates (e.g. rabies) increase

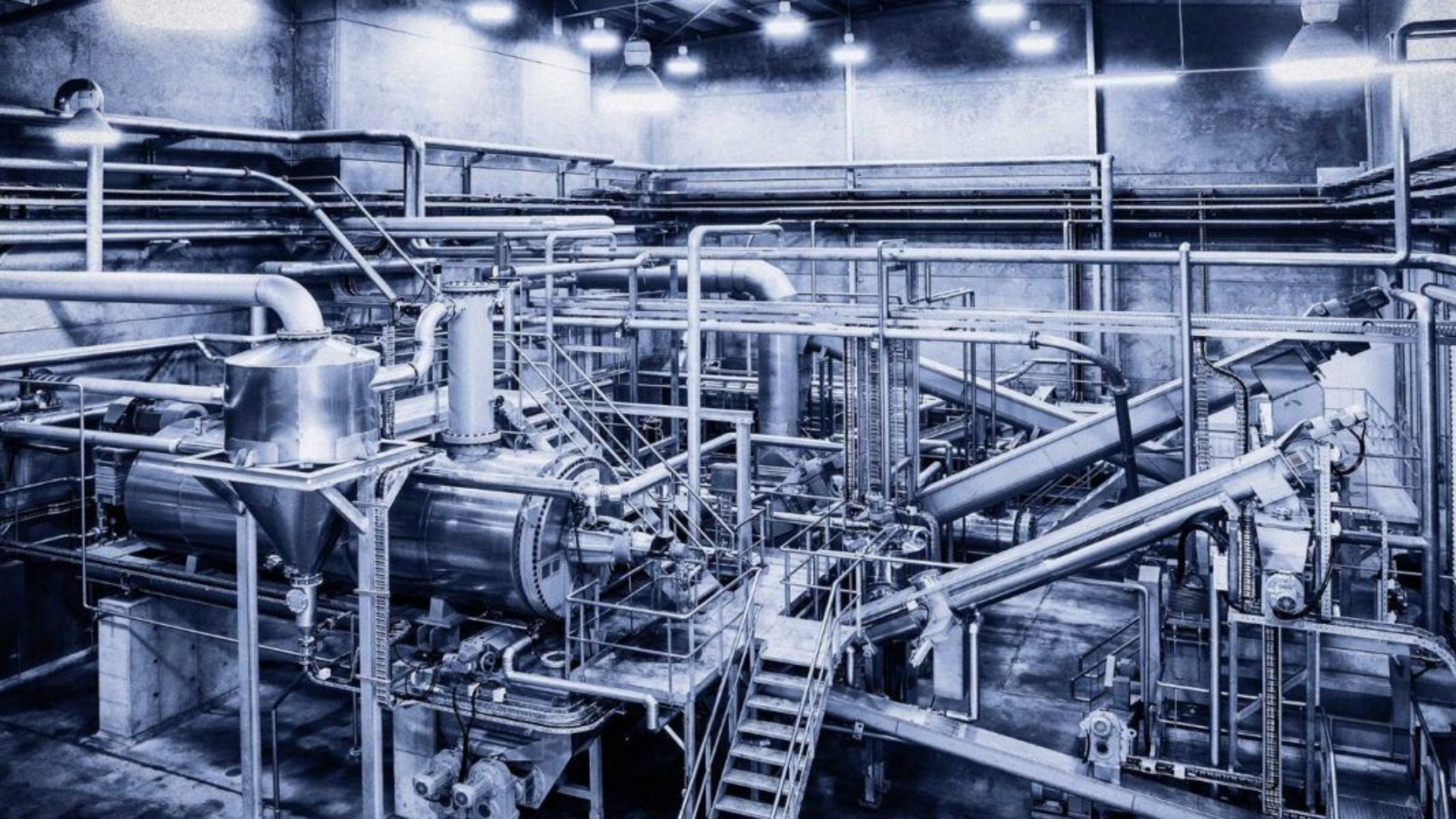


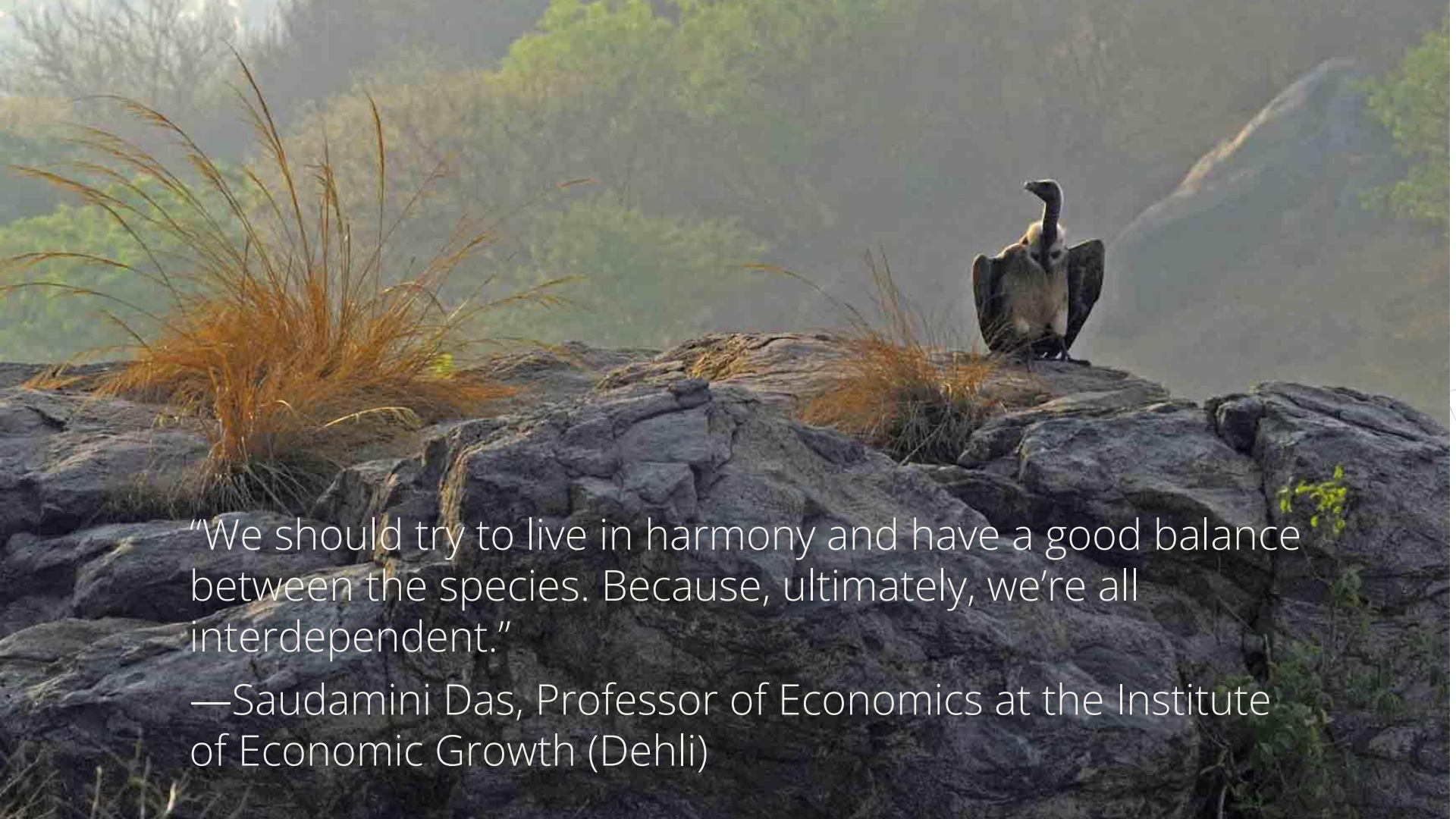
Increased human disease & death



Water quality decreases

Livestock carcasses build up





"We should try to live in harmony and have a good balance between the species. Because, ultimately, we're all interdependent."

—Saudamini Das, Professor of Economics at the Institute of Economic Growth (Dehli)

Courtney Nash

Head of The VOID

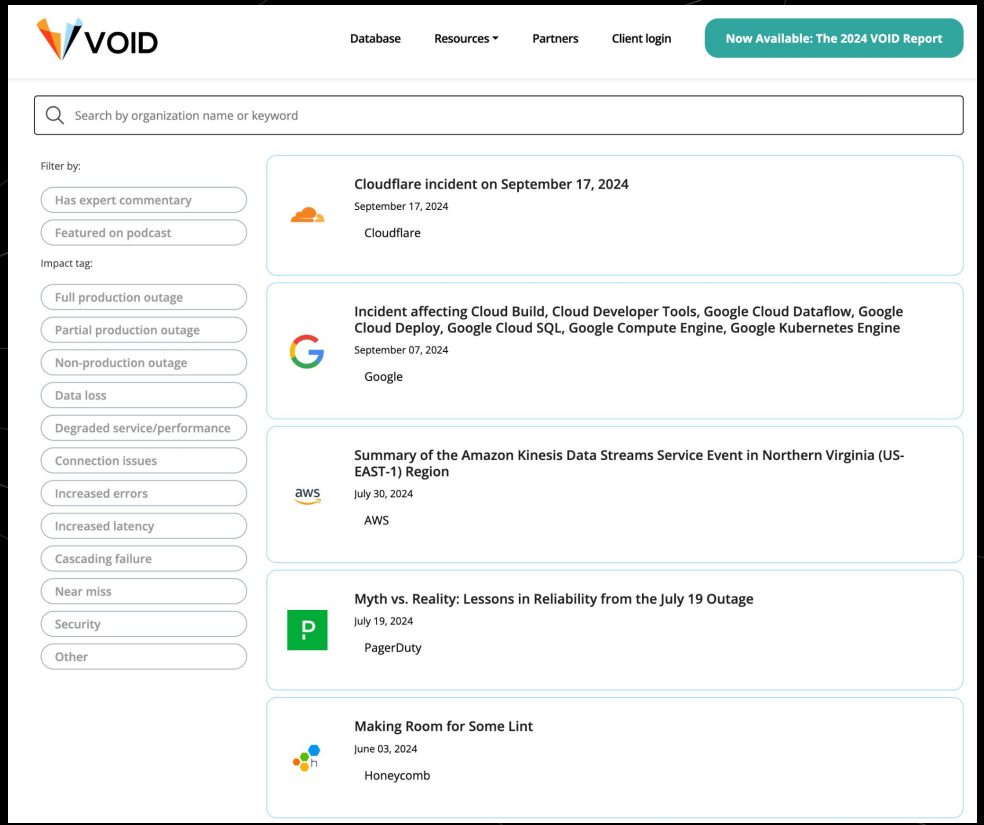
amazon Microsoft O'REILLY® fastly VERICA

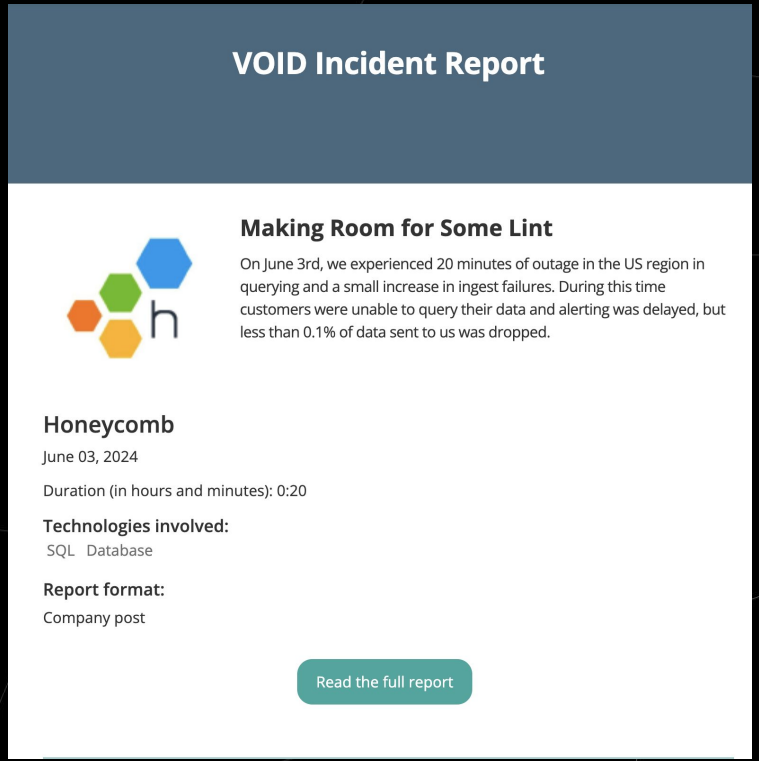
About The VOID

https://www.thevoid.community

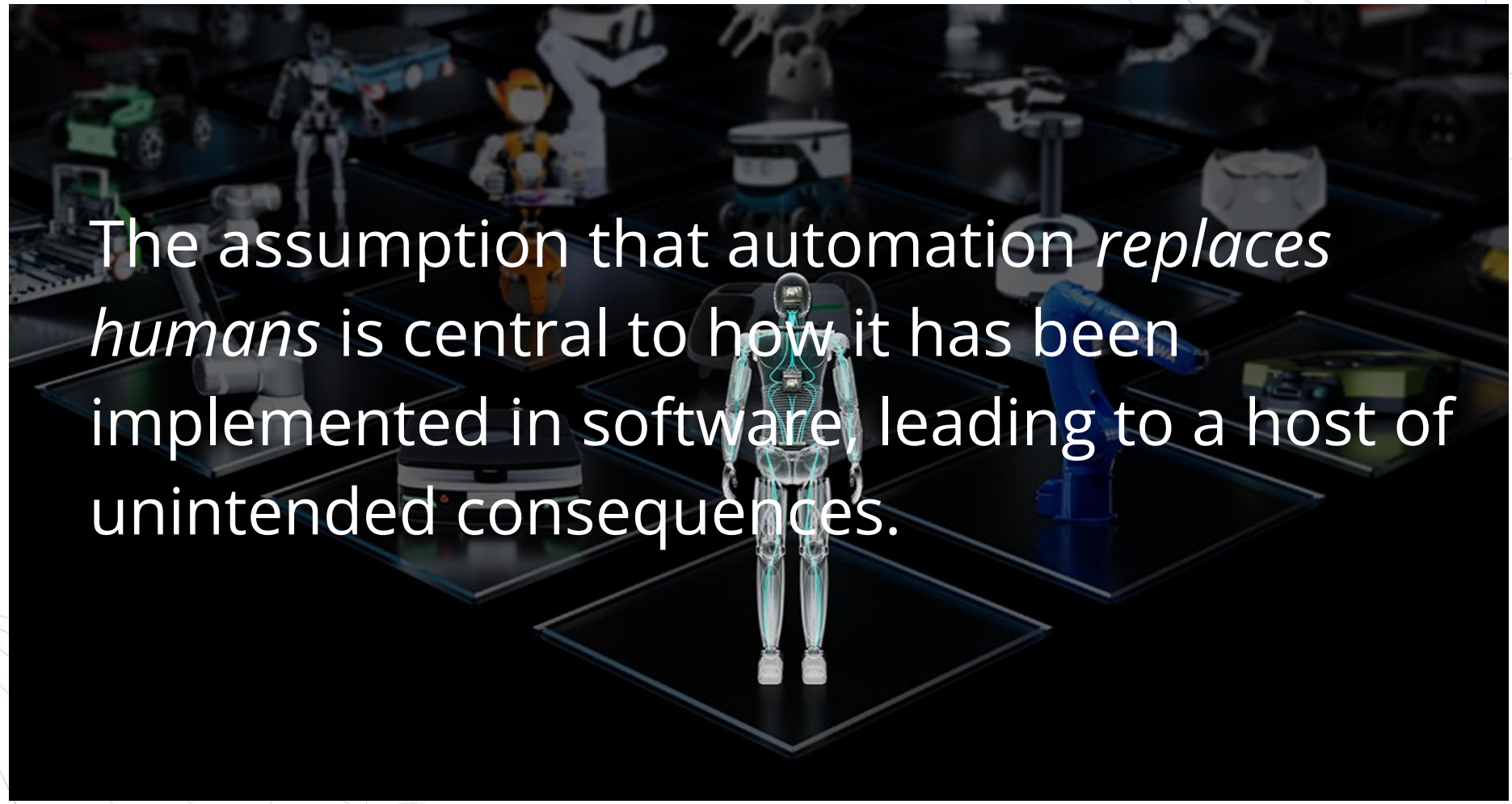
What's In The VOID?

https://www.thevoid.community





The Unexpected Consequences of Automation in Software

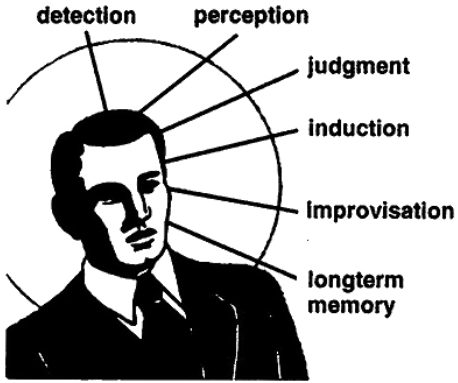

The assumption that automation *replaces humans* is central to how it has been implemented in software, leading to a host of unintended consequences.

But First: Research From Other Domains

Functional Allocation & The Substitution Myth

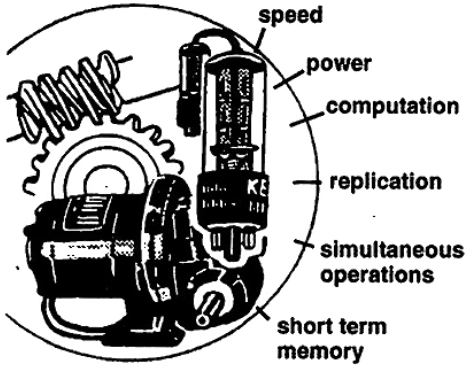
(AKA The Fitts List)

HUMANS SURPASS MACHINES IN THE:



- Ability to detect small amounts of visual or acoustic energy
- Ability to perceive patterns of light or sound
- Ability to improvise and use flexible procedures
- Ability to store very large amounts of information for long periods and to recall relevant facts at the appropriate time
- Ability to reason inductively
- Ability to exercise judgment

MACHINES SURPASS HUMANS IN THE:



- Ability to respond quickly to control signals, and to apply great force smoothly and precisely
- Ability to perform repetitive, routine tasks
- Ability to store information briefly and then to erase it completely
- Ability to reason deductively, including computational ability
- Ability to handle highly complex operations, i.e., to do many different things at once.

The Substitution Myth in Automation

Fixed Strengths and Weaknesses

Separate Tasks

Dekker & Woods, 2002

1. Designers of automation tend to **imagine the desired outcomes of automation** (e.g. lower workload, higher accuracy) and that only those desired outcomes will occur (see also Norman 1990).
2. Automation does not have access to all real-world parameters for accurate problem solving in all contexts, and may in fact make it **harder for humans to directly impact the system** in the ways they want.
3. Allocating aspects of the system to automation **creates new categories or functions that humans must take on**, such as figuring out where to find information about what the automation is actually doing.
4. Automation does not necessarily replace human weaknesses. It often **creates new human weaknesses** or requires the development of new, unanticipated strengths.

Ironies of Automation

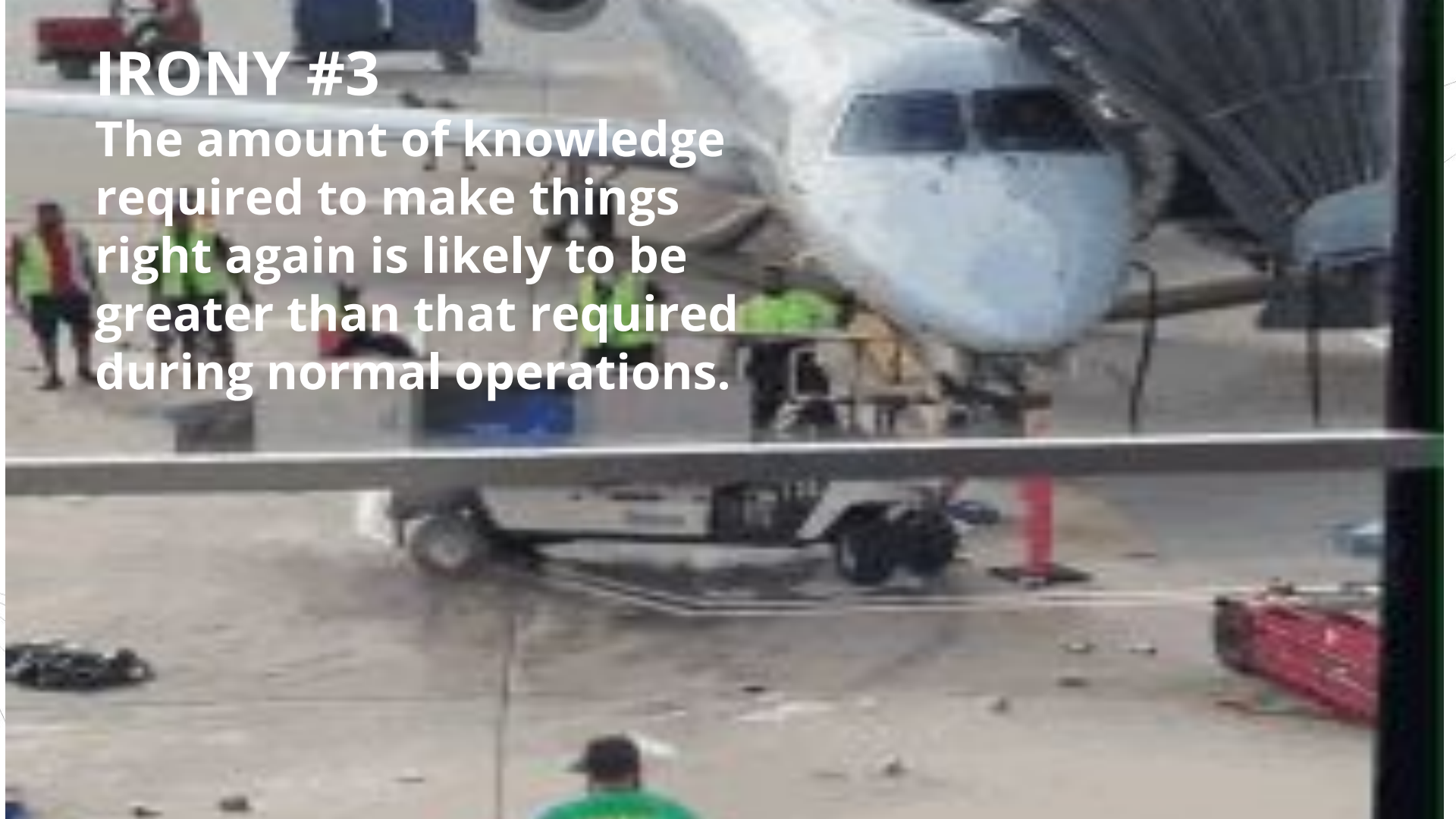
Bainbridge, L. (1983). The Ironies of Automation. *Automatica*, 19, 775-779. (Conference proceedings).

IRONY #1

Humans design the automation and then also deal with its unanticipated, often negative consequences.

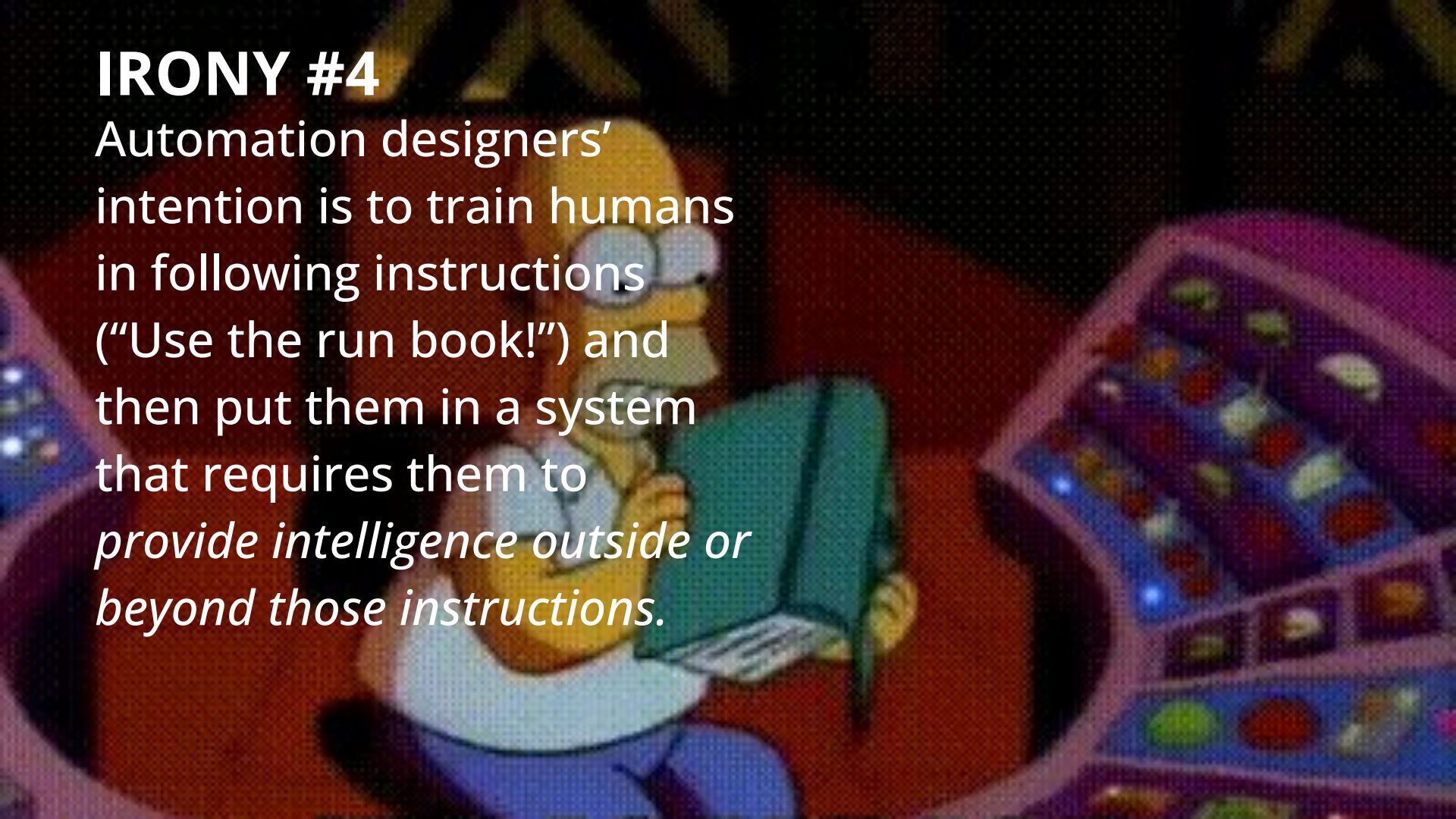
IRONY #2

Human operators have to monitor that the automation is working properly.



IRONY #3

The amount of knowledge required to make things right again is likely to be greater than that required during normal operations.



IRONY #4

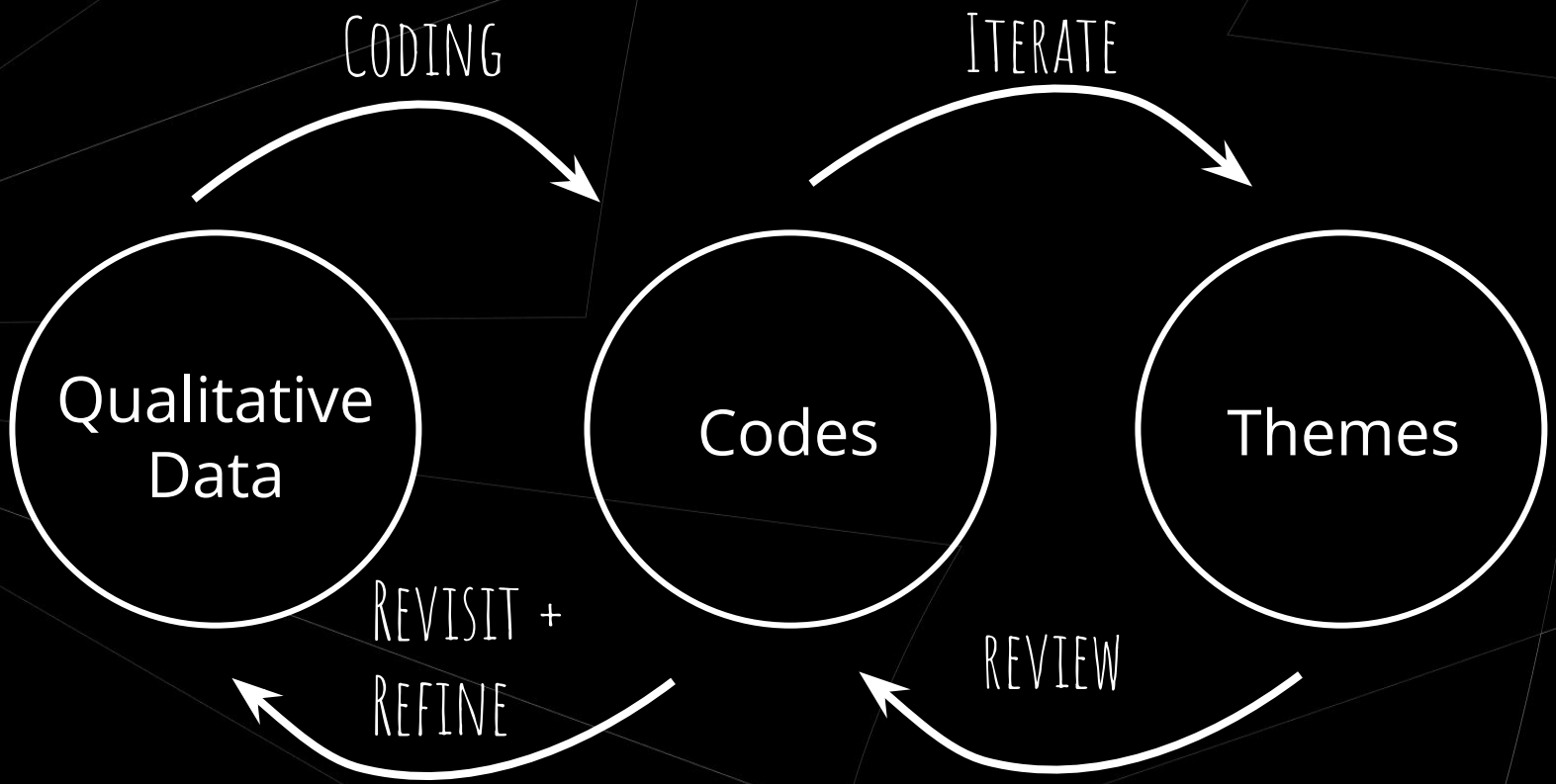
Automation designers' intention is to train humans in following instructions ("Use the run book!") and then put them in a system that requires them to *provide intelligence outside or beyond those instructions.*

“...The more advanced a control system is, the more crucial may be the contribution of the human operator.”

—Bainbridge, 1983

Research From The VOID

Thematic Analysis



Methodology: Keyword search

VOID

10k+

Automation
Load balancing
CI/CD
Self-heal/ing

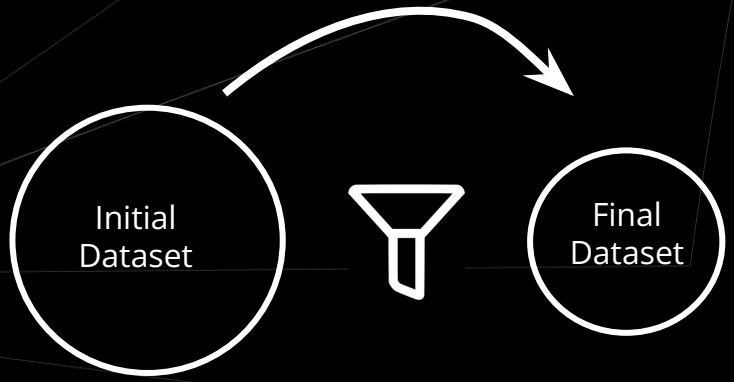
Automated
Load balancer
Retry storm
Alert/ing/ed

>

Dataset

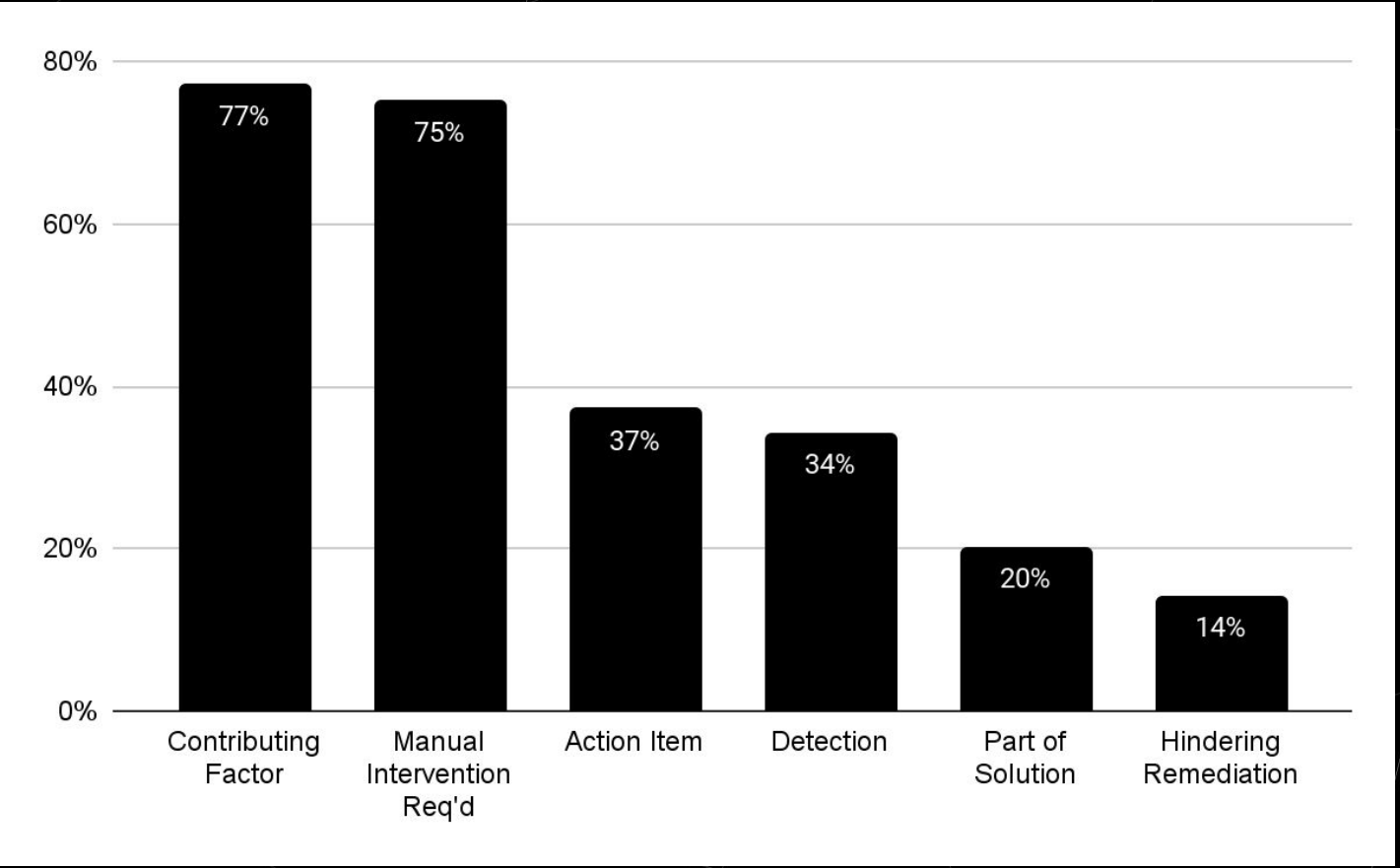
459

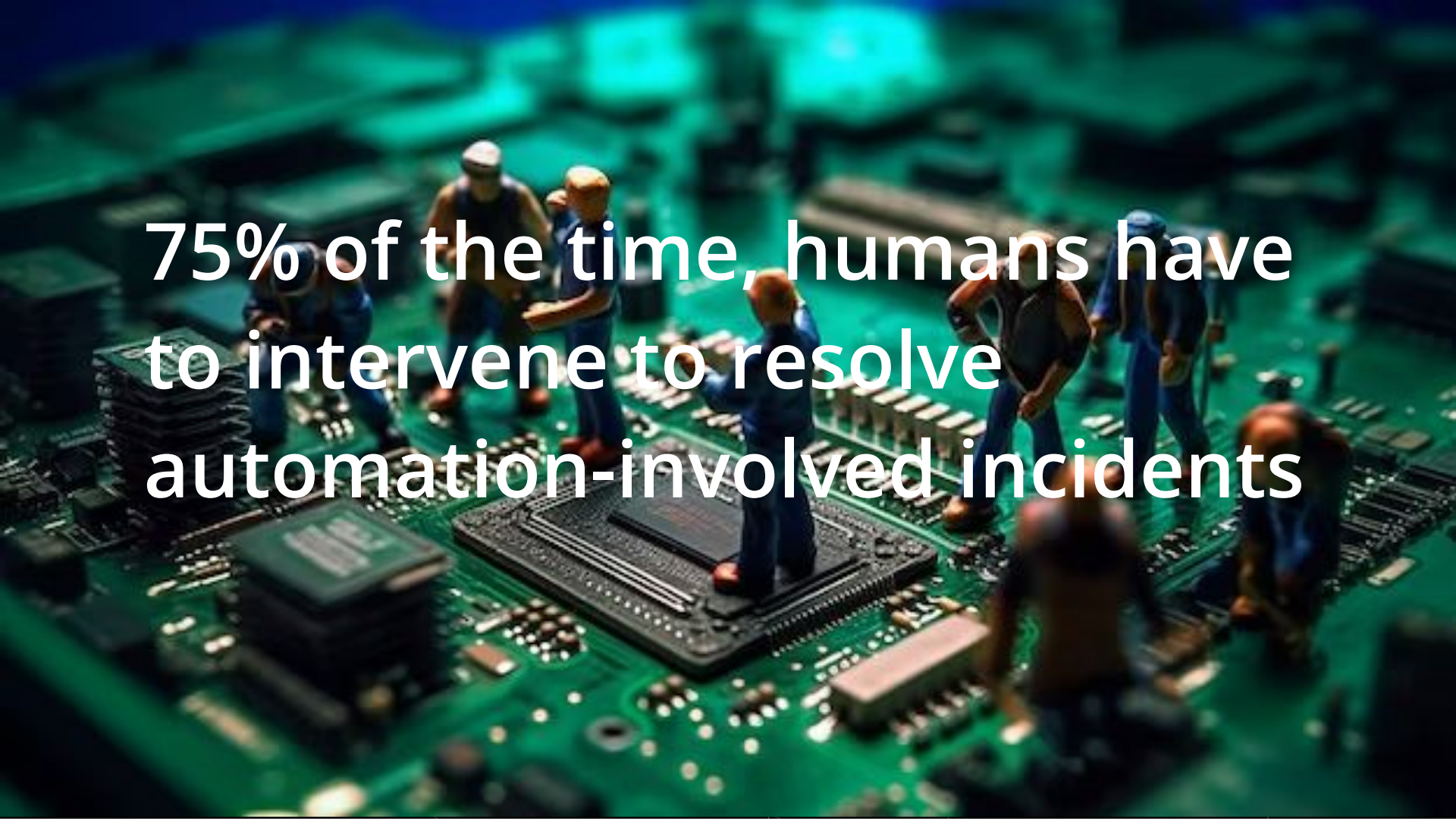
Methodology: Initial Coding



- Detection
- Contributing Factor
- Hindering Remediation
- Involved in Solution
- Manual Intervention
- Action Item

Quantitative Results




75% of the time, humans have to intervene to resolve automation-involved incidents

Automation Themes

I. Automation Plays Multiple Roles in Incidents

II. Automation Can Unexpectedly Make Things Worse

III. Human Intervention Remains Essential to Resolve Issues

Better Automation Through Joint Cognitive Systems

Automation: Expectations vs. Reality

An “un-Fitts” List for Joint Cognitive Systems

MACHINES	
Are constrained in that	**Need people to**
Sensitivity to context is low and is ontologically limited	Keep them aligned to context
Adaptability to change is low and recognition of anomalies is ontologically limited	Keep them stable given the variability and change inherent in the world
Adaptability to change is low and is ontologically limited	Repair their ontologies
They are not “aware” of the fact that the model of the world is itself in the world	Keep their model aligned with the world

PEOPLE	
Are not limited in that	**Yet they create machines to**
Sensitivity to context is high and is knowledge- and attention-driven	Help them stay informed of ongoing events
Adaptability to change is high and is driven by the recognition of anomaly	Help them align and repair their perceptions because they rely on mediated stimuli
Adaptability to change is high and is goal-driven	Effect positive change following situational change
They are aware of the fact that the model of the world is itself in the world	Computationally create their models of the world

10 Aspects of Joint Cognitive Systems

1. To be a team player, an intelligent agent must fulfill the requirements of a Basic Compact to engage in common-grounding activities.
2. To be an effective team player, intelligent agents must be able to **adequately model the other participants' intentions and actions** vis-a-vis the joint activity's state and evolution—for example, are they having trouble?
3. Human-agent team members must be **mutually predictable**.
4. Agents must be **directable**.
5. Agents must be able to **make pertinent aspects of their status and intentions obvious** to their teammates.
6. Agents must be able to **observe and interpret pertinent signals of status and intentions**.
7. Agents must be able to **engage in goal negotiation**.
8. Support technologies for planning and autonomy must **enable a collaborative approach**.
9. Agents must be able to **participate in managing attention**.
10. All team members must help **control the costs of coordinated activity**.

Questions?