Contestable AI & Civic Co-Design

Kars Alfrink
TU Delft
contestable.ai

Reinventing the City
17 February 2022

Credit: Bloomberg

Problem

Automated decision-making can harm people's basic human rights to autonomy and dignity.

Solutions

Ex-ante: Participatory & value-sensitive design

Ex-post: Explanations, appeal procedures

"When change is easy, the need for it **cannot be foreseen**; when the need for change is apparent, **change has become expensive**, difficult, and time-consuming."

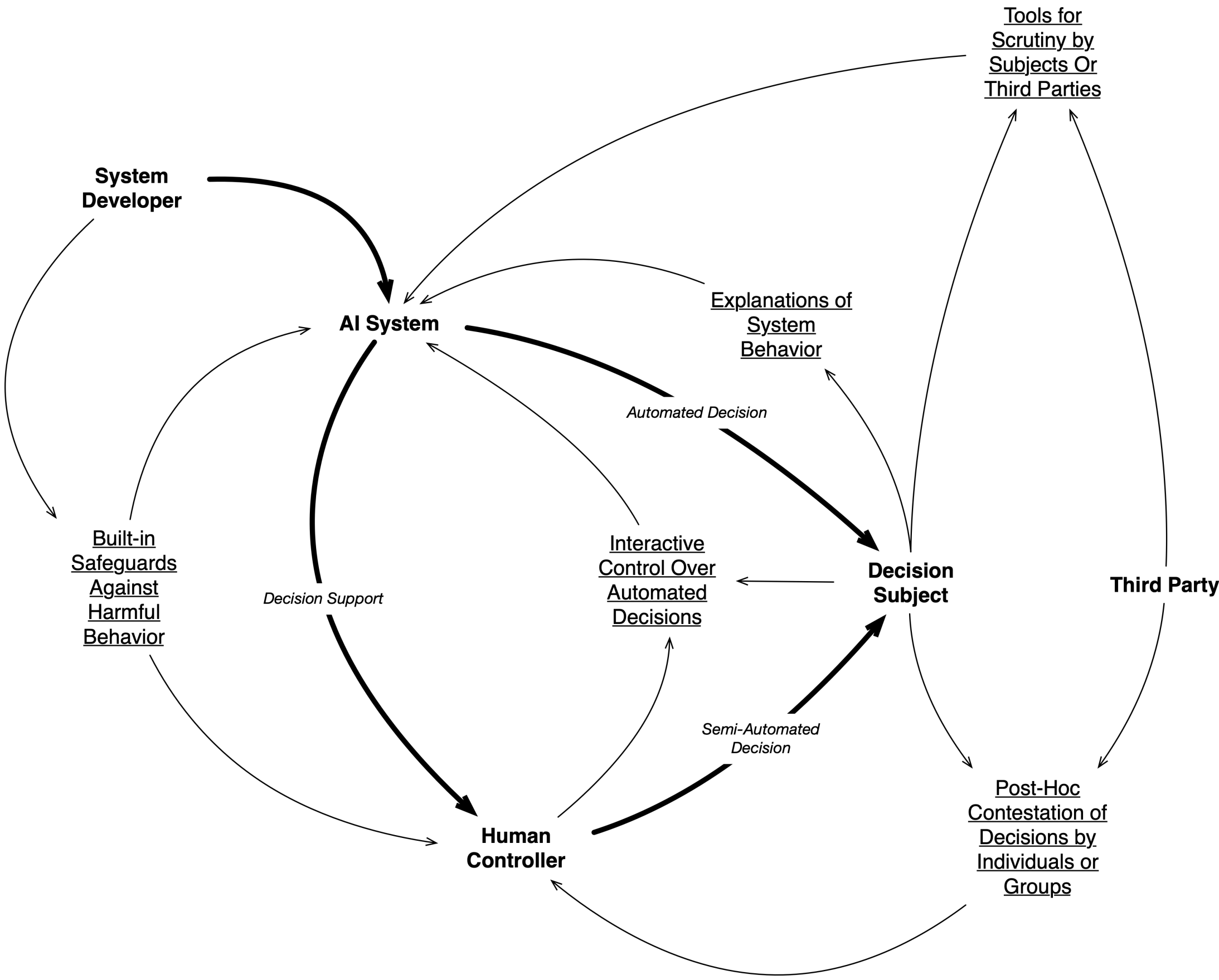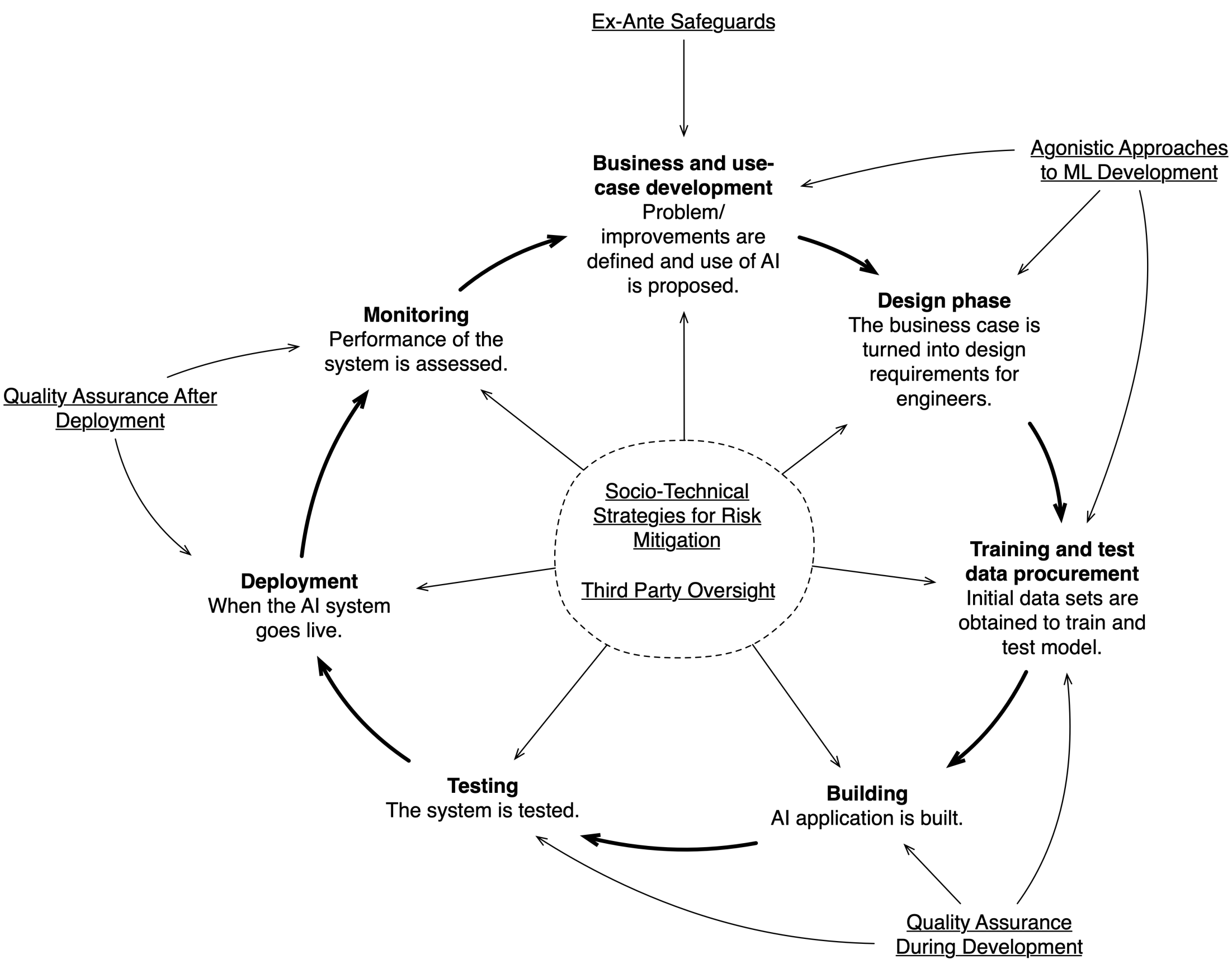Collingridge, D. (1980). The social control of technology. Frances Pinter.

Contestable AI

AI that is **open and responsive to human intervention**, **throughout a system's lifecycle**, establishing a **procedural relationship** between decision subjects and system controllers.

Features   Practices

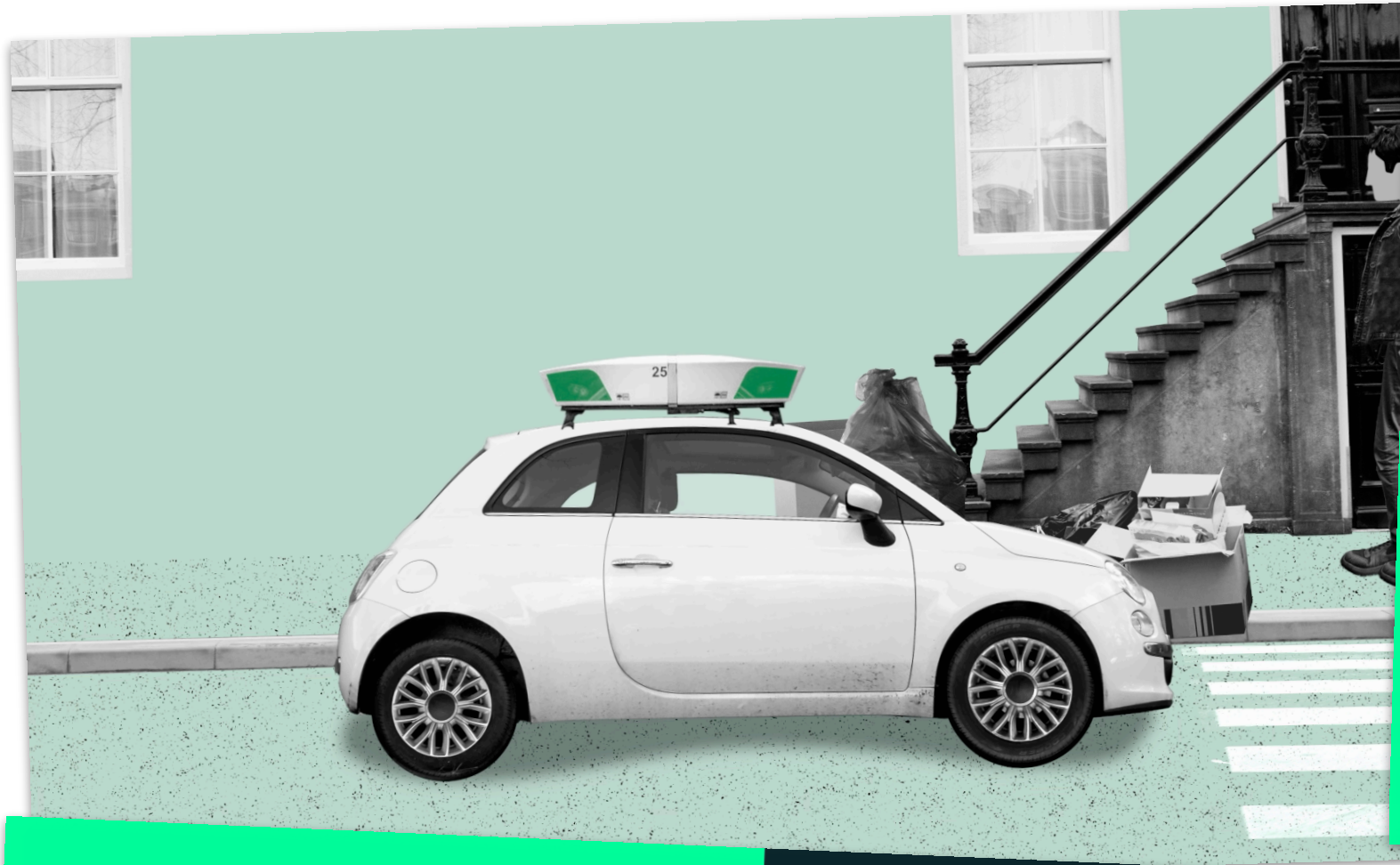Alfrink, K., Keller, I., Kortuem, G., & Doorn, N. (2021). Contestable AI by Design: Towards A Framework. Manuscript Submitted for Publication.

Work in progress

Contestable Camera Cars

A speculative design exploration of public AI systems that are responsive to value change.

camera car

connected

trash
60% sure

heat map

vehicle	destination
🚛 truck 03	→ location a
🚛 truck 12	→ location b
🚛 truck 17	→ location a
🚛 truck 09
🚛 truck 02
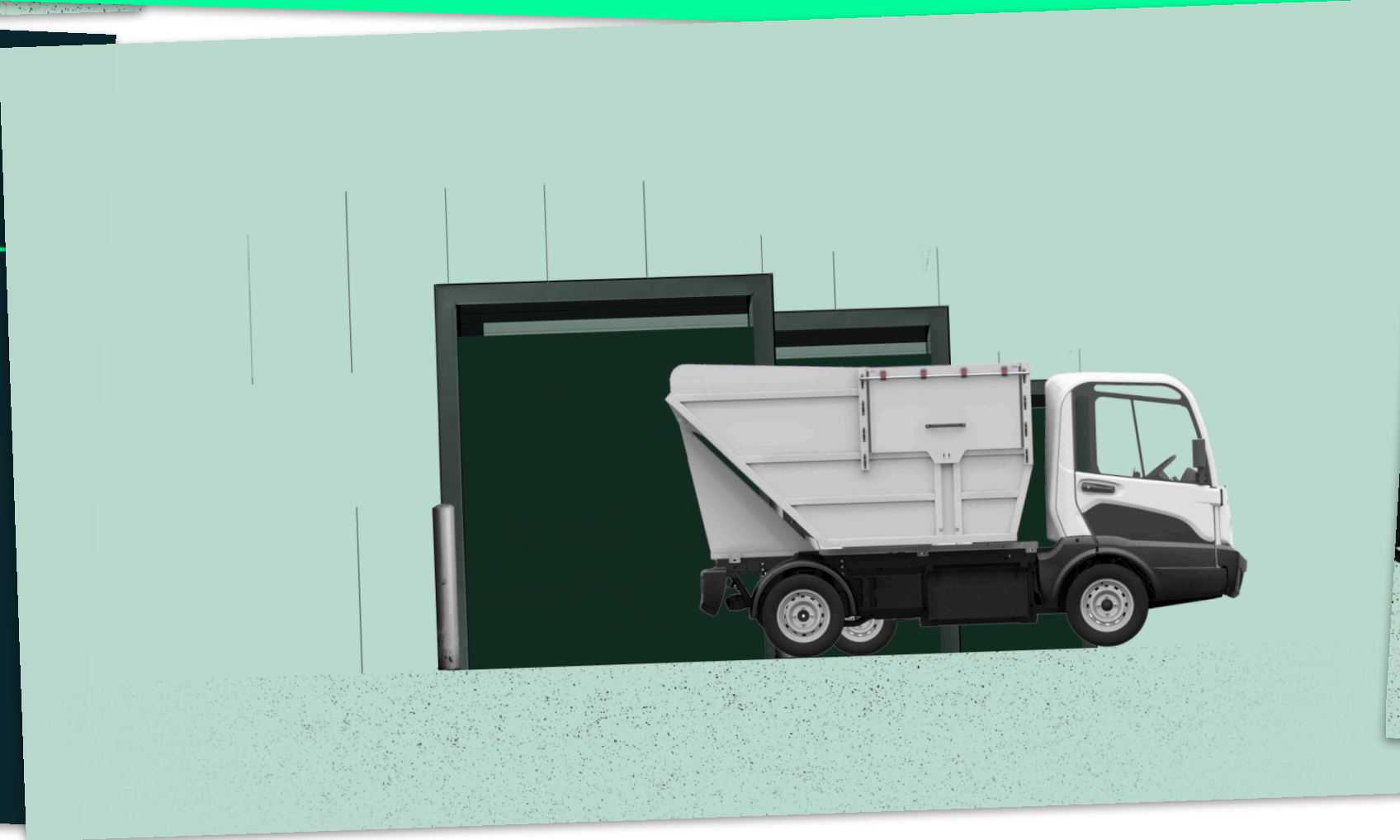
Takeaways

Avoid **resolving disputes up front** using compromise or consensus-seeking.

Set up **procedural, agonistic mechanisms** to identify and resolve them.

Leverage **conflict** for **continuous improvement**.

Thank you!

Kars Alfrink
TU Delft
contestable.ai

CONTESTABLE AI
Design research for human intervention in AI systems

To ensure artificial intelligen
to autonomy and dignity, the
during development and afte

This PhD research project ai
design of mechanisms that (
individual automated decisio
people to collectively contes
systems, particularly as it pe

AI system transparency, fairr
that can be solved by purely
require a careful consideratio
together, and should take loc
challenges that design resea

https://contestable.ai