HUMAN AND AI

Dr. Jurriaan van Diggelen

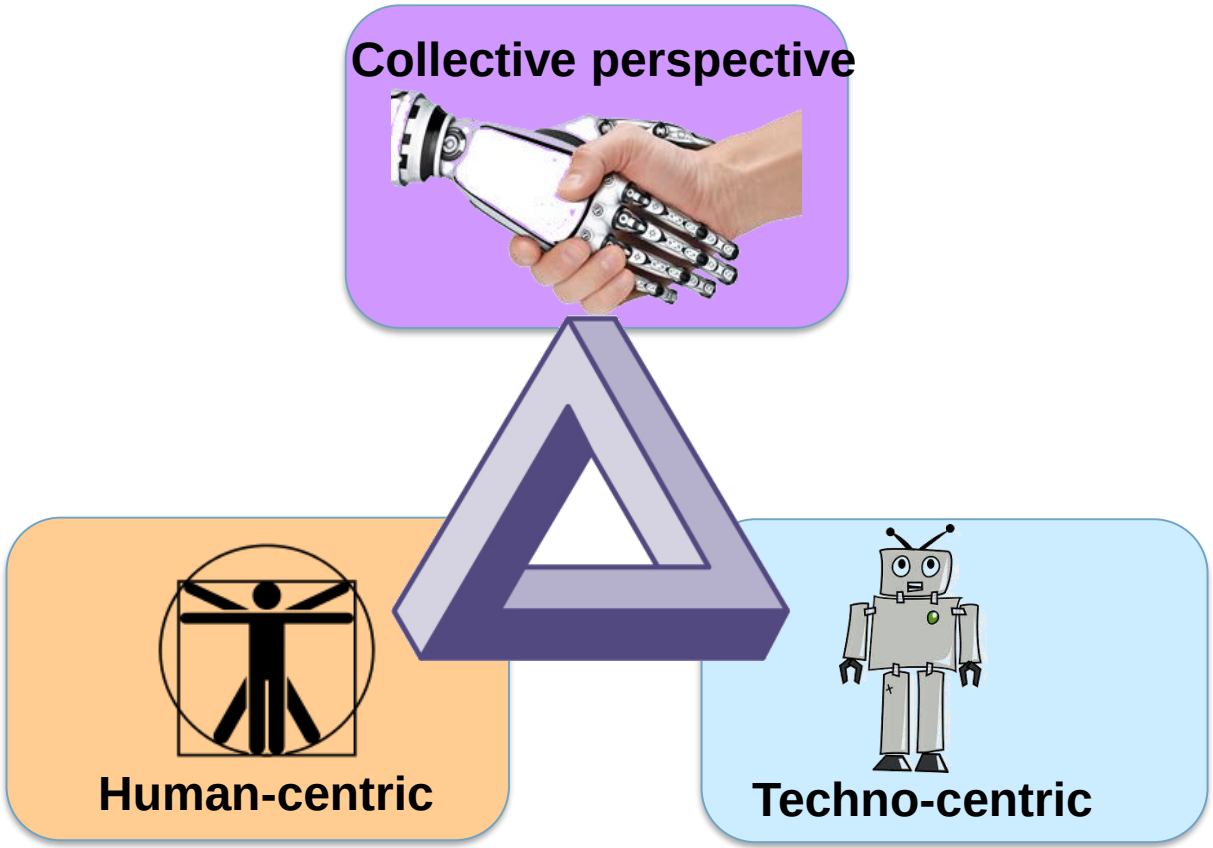
28 november 2019

Human Factors NL Jaarcongres 2019

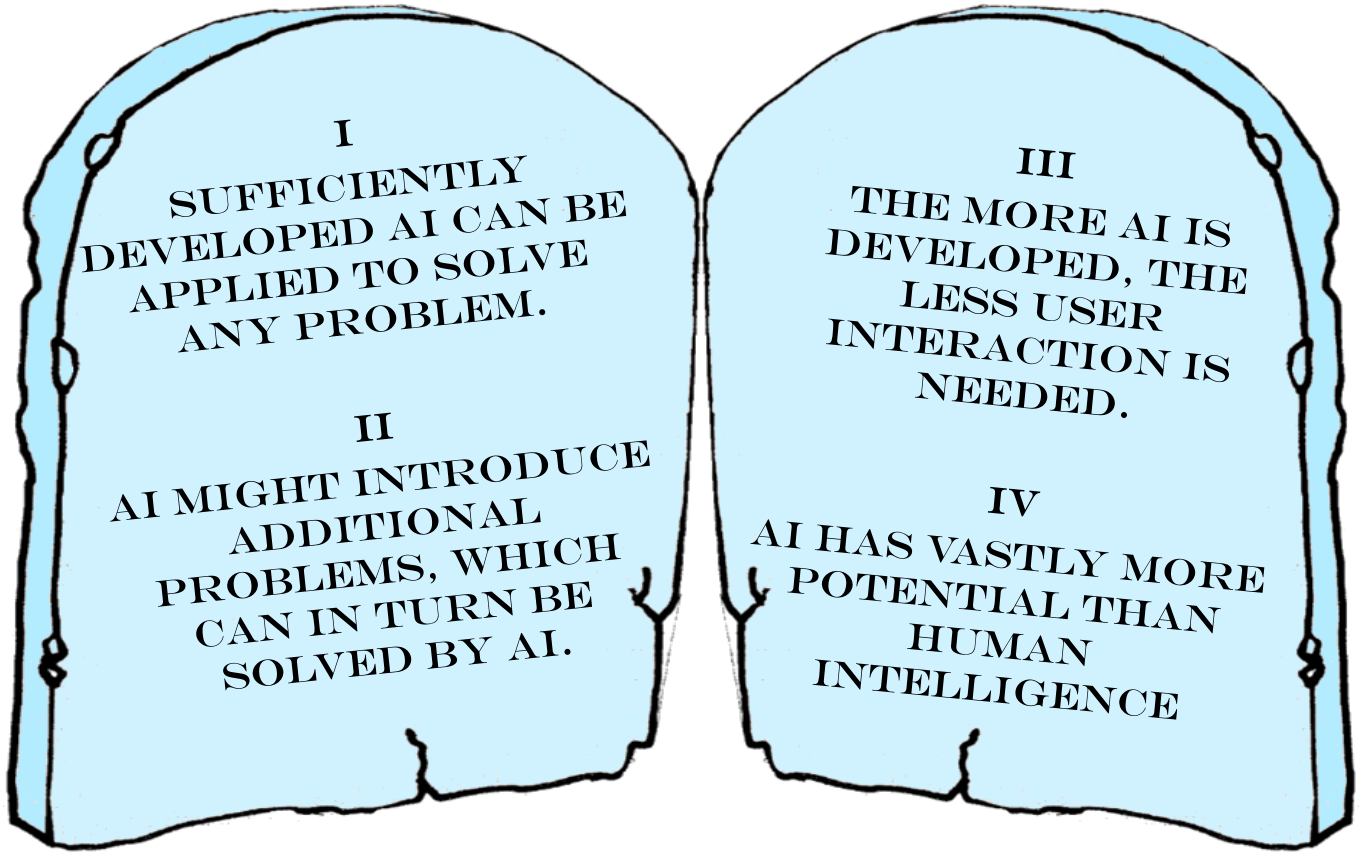
TNO innovation for life

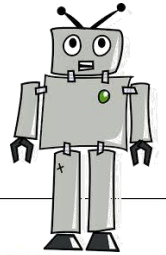
Part I
Three perspectives on AI

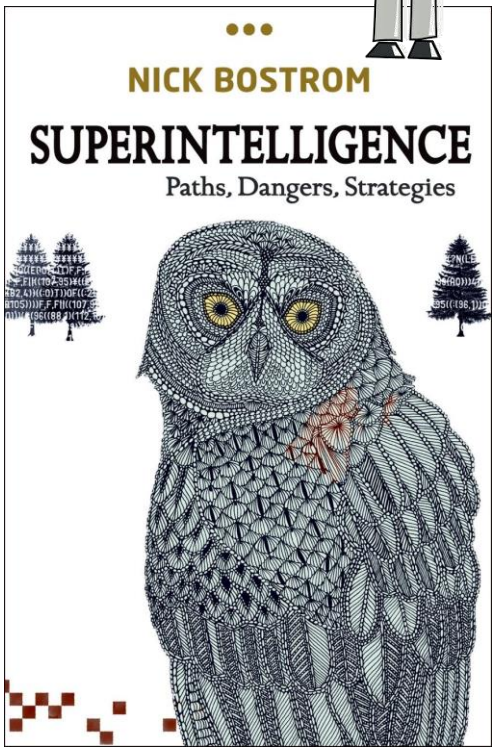
3 PERSPECTIVES ON ARTIFICIAL INTELLIGENCE



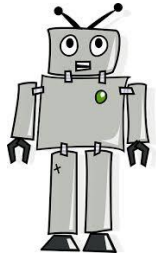
TECHNO-CENTRISM

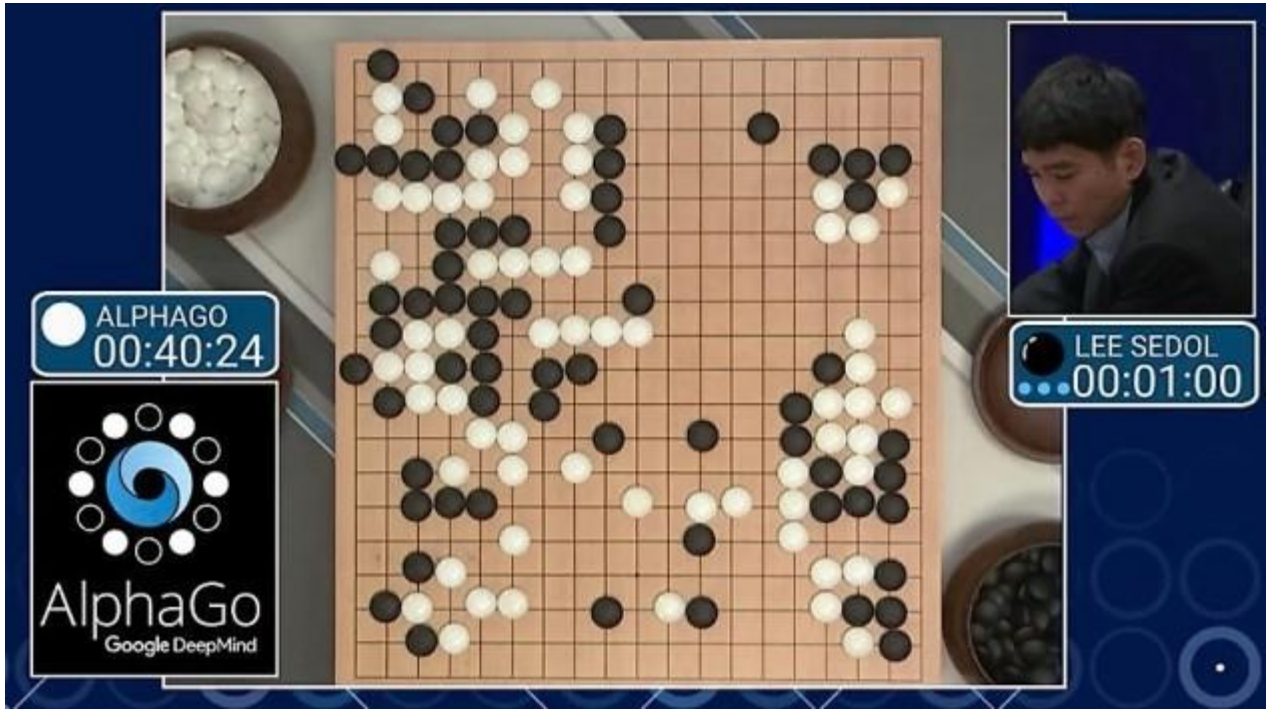




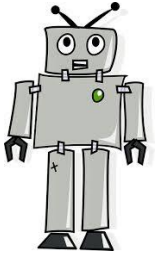


ALPHA (GO) ZERO



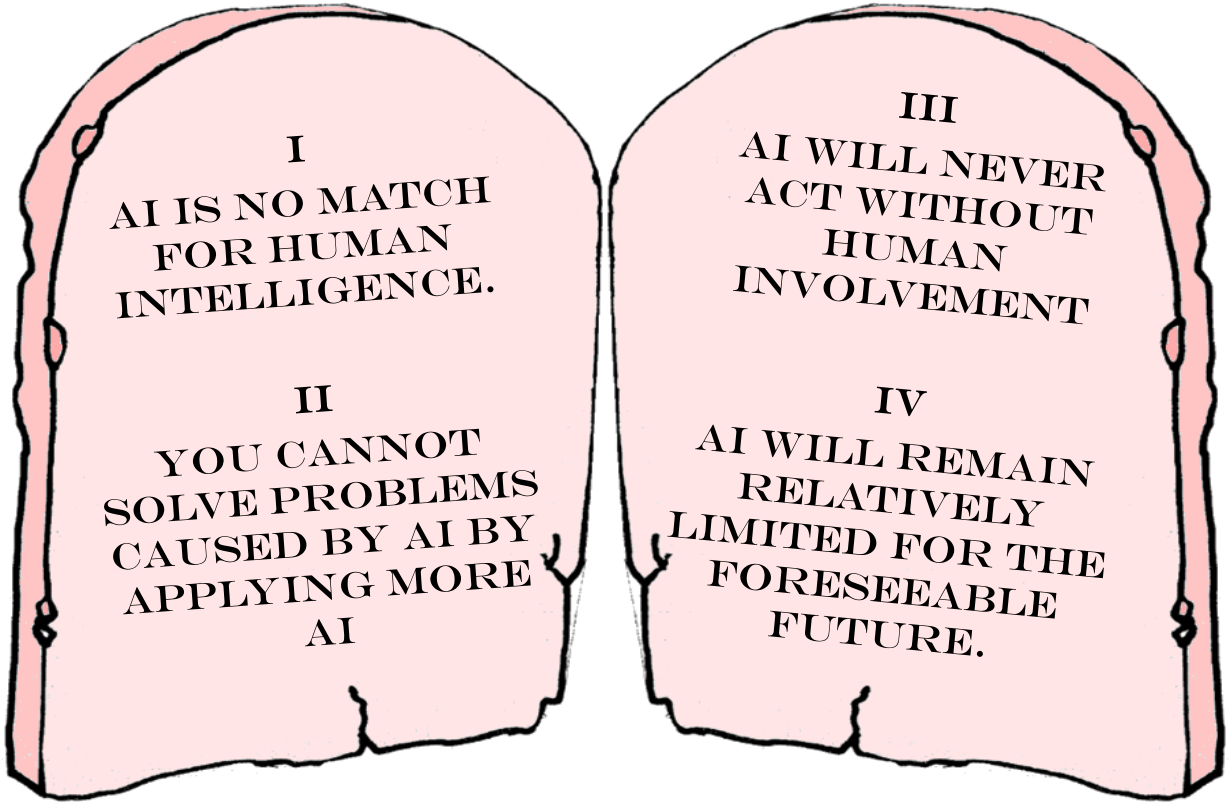


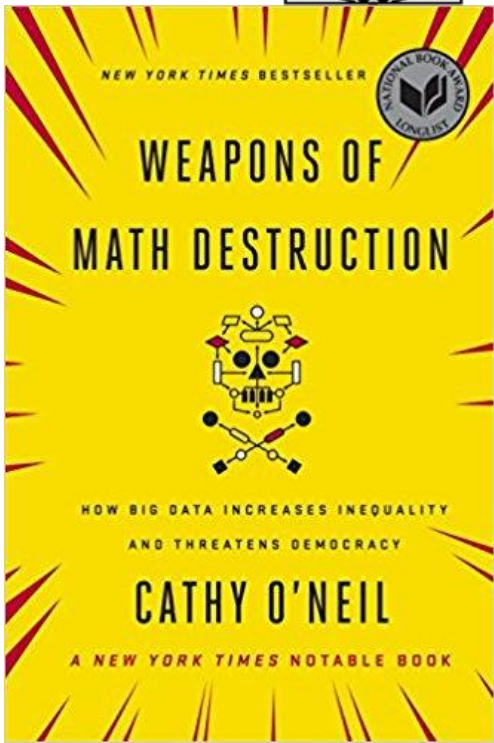
THE ROLE OF THE HUMAN



- **Problem:** How to maintain human control over Artificial Super Intelligence.
- **Solution:** Program human values into AI system

HUMAN CENTRISM





PROPERTIES OF A WMD

- **Non-transparant**: It is unclear how AI arrives at its conclusions.
- **Scale**: The decisions made by AI affect large groups of people.
- **Damage:** The AI brings damage to large groups of people.

TEACHER ASSESSMENT TOOL

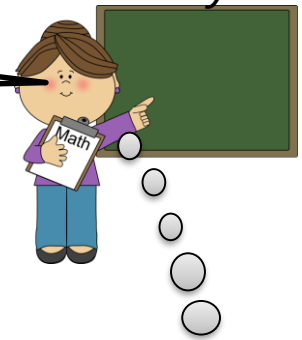
We must turn around underperforming schools in Washington D.C.

M50
MATHEMATICA
Policy Research

We develop an objective and accurate model IMPACT to assess a teacher's performance

Sarah Wysocki

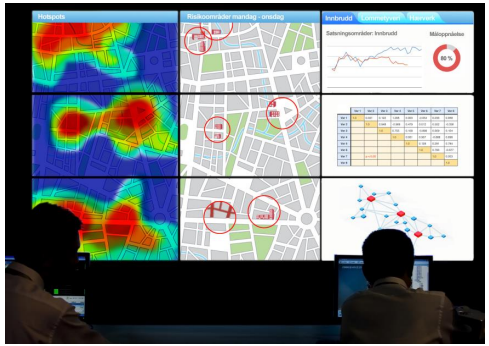
Along with 205 other teachers with a low IMPACT score, I got fired. Why?



It's a complex algorithm you won't understand. Furthermore, it's corporate secret.

Many of my students came from a different school where they tampered test scores. They started scoring less in my tests…

OTHER EXAMPLES OF WMD'S



Predictive policing

BUSINESS NEWS OCTOBER 10, 2018 / 5:12 AM / 6 MONTHS AGO

REUTERS

Amazon scraps secret AI recruiting tool that showed bias against women

Scan CV's

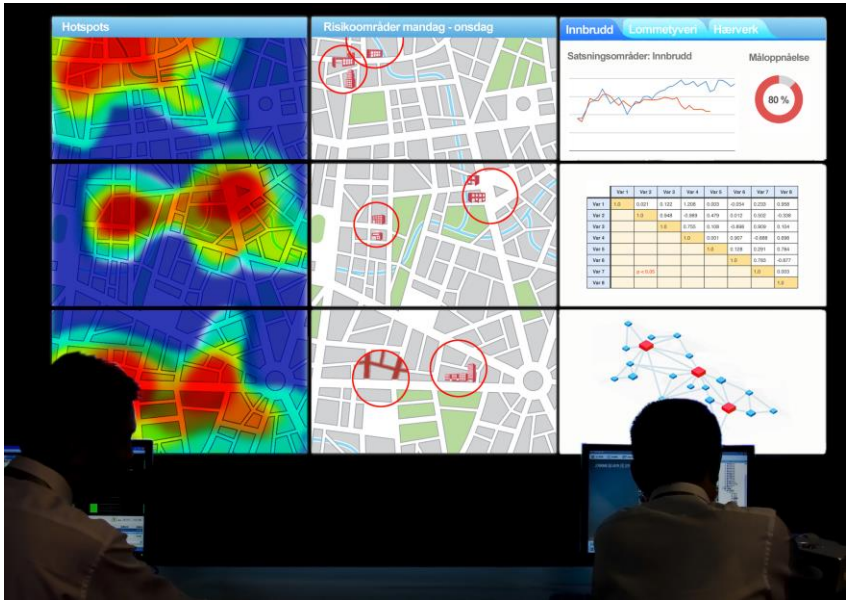
Political campaigning

Assess creditworthiness

Predict chance of recidivism

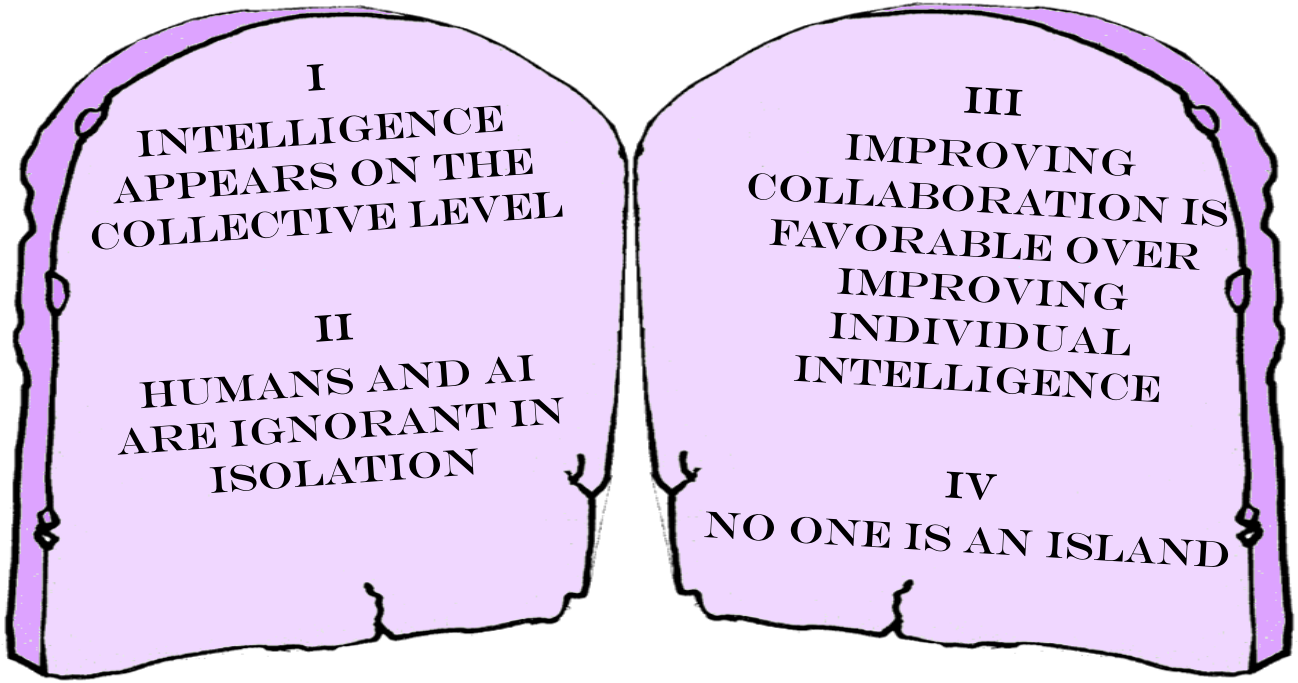
Calculate insurance premium

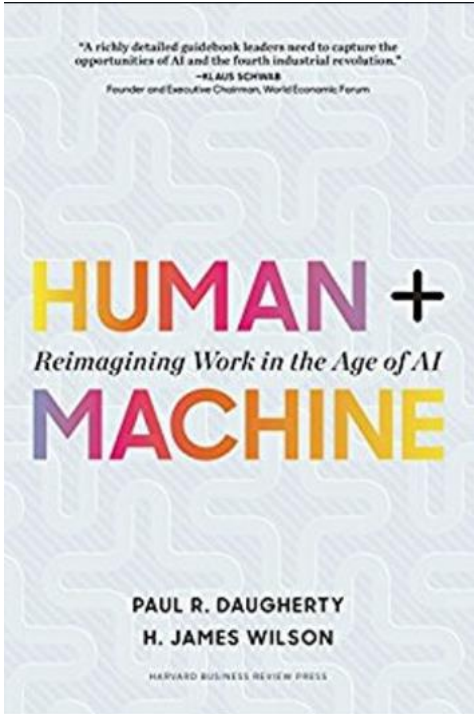
THE ROLE OF THE HUMAN



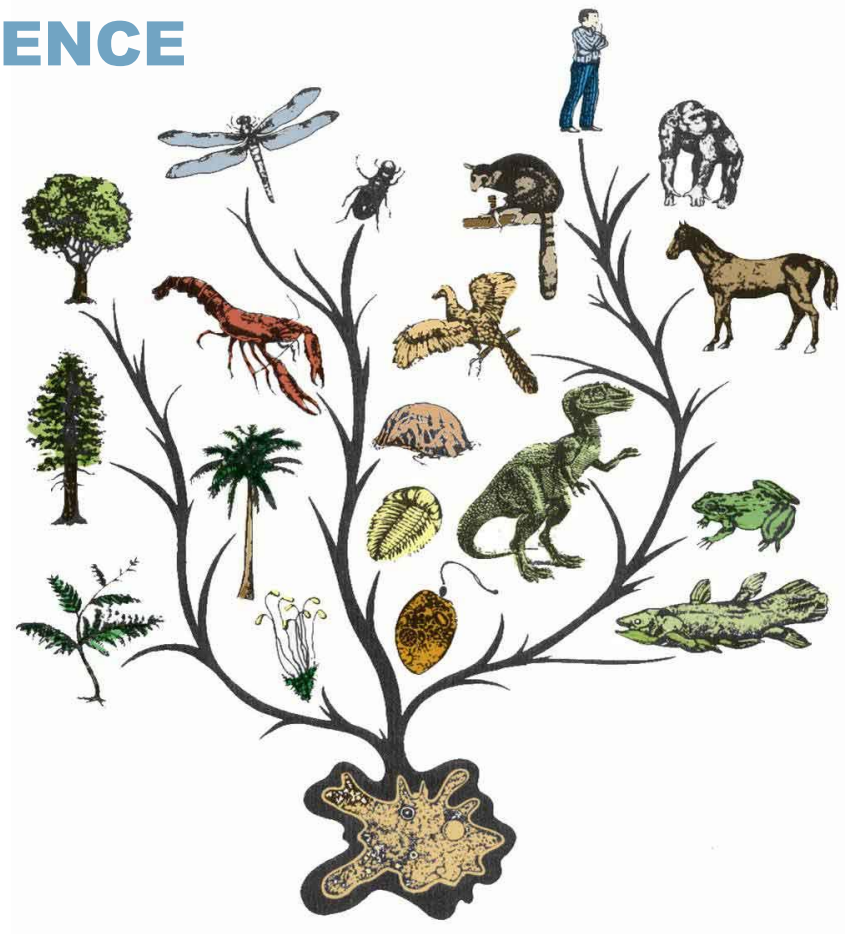
- **Problem:** Oversimplified AI models are granted too much control.
- **Solution:** Apply AI sparingly.

COLLECTIVE INTELLIGENCE

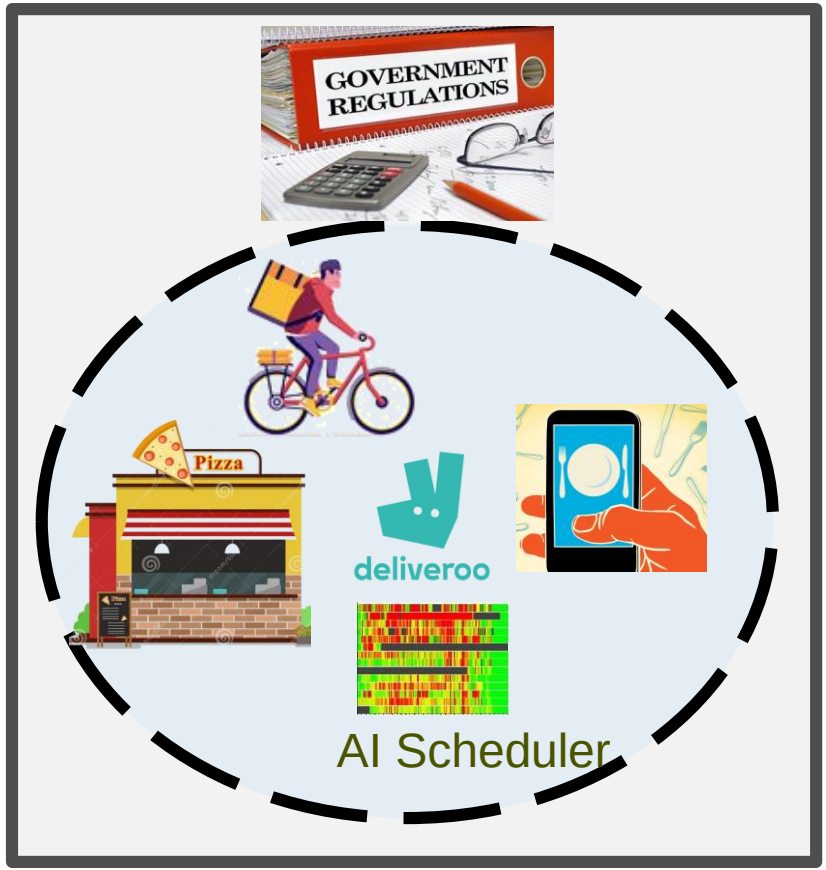




SOCIAL INTELLIGENCE



SOCIAL AI IS ESSENTIAL


GOVERNMENT
REGULATIONS
Pizza
deliveroo
AI Scheduler

THE ROLE OF THE HUMAN

Problem: Integrating AI into teams, organisations, and society inevitably disturbs the equilibrium between autonomy and control.

Solution: Detect and redirect undesirable developments.

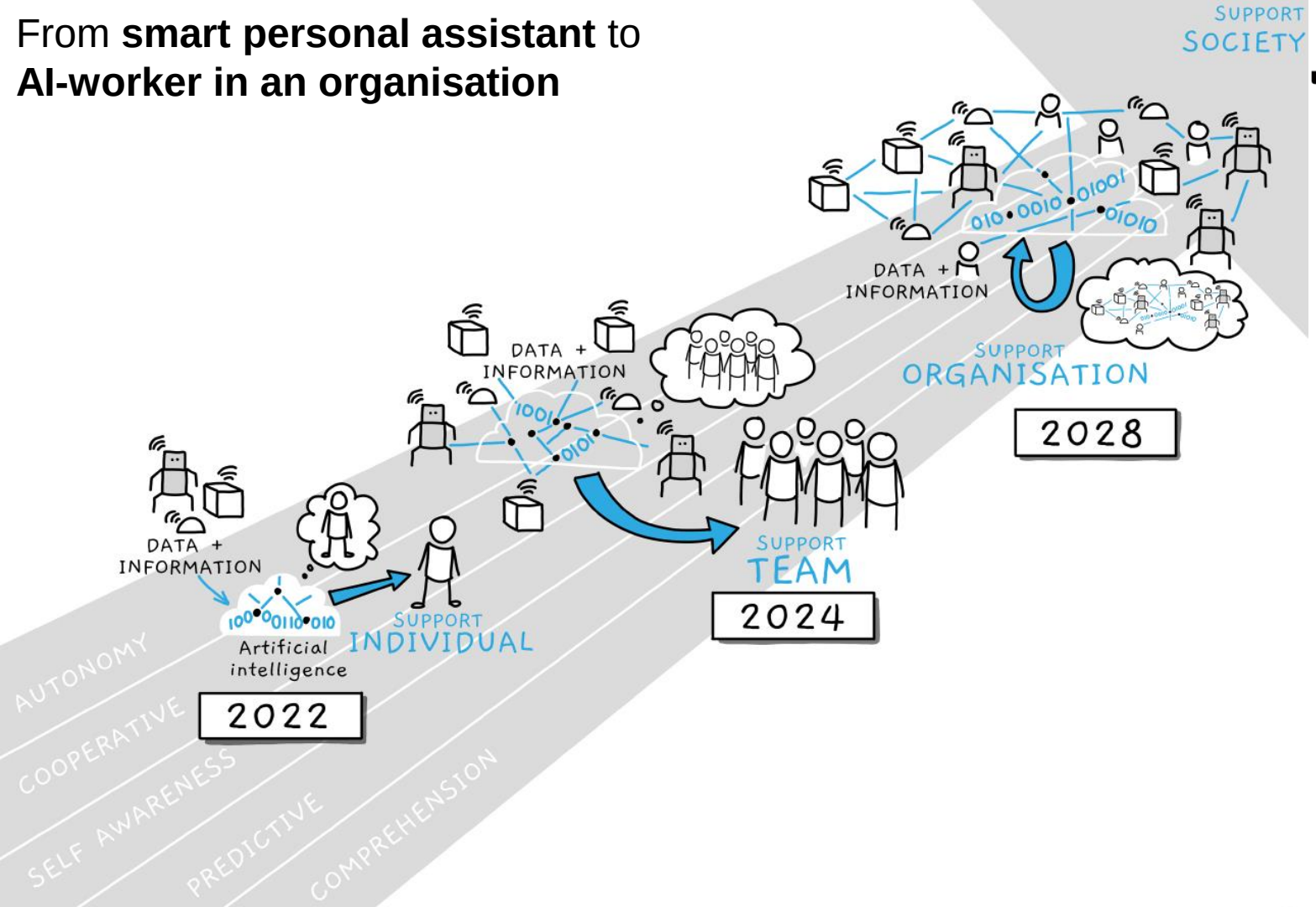
EXAMPLE RADIOLOGY

- Deep learning networks achieve super human performance, on segmentation and identification of malicious tissue in X-ray images.

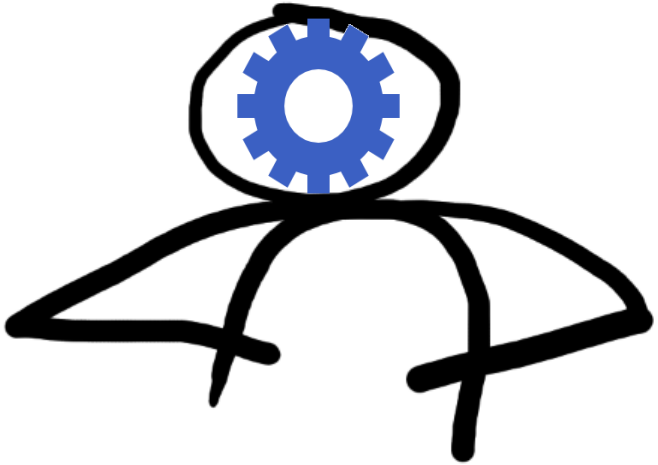
- AI is expected to revolutionize radiology
  - **AI-replacement**: Some tasks are completely taken over by AI. E.g. visual interpretation of radiology images. Results in deskilling of radiologists.
  - **AI-augmentation**: For some tasks, the AI system augments the human. E.g. planning a patient treatment. Requires the human to maintain expertise and acquire additional expertise on how to use the AI support system.
  - **AI-maintenance:** These are tasks that are added to the radiology workflow that did not exist before. Requires a whole new set of skills. Examples are: (re-)training the AI system.

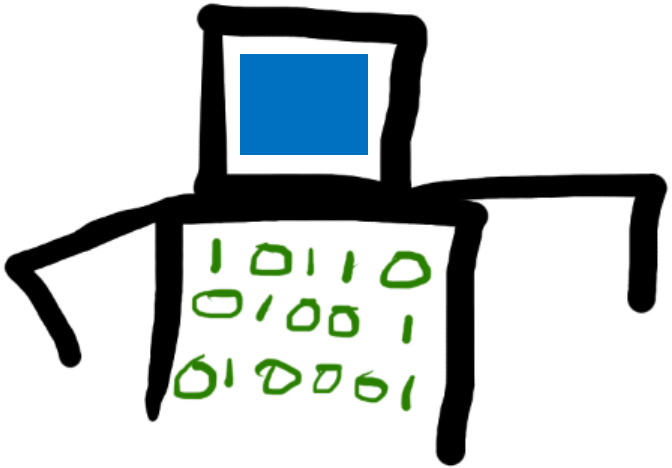
Part II
Team Design Patterns

From **smart personal assistant** to **AI-worker in an organisation**

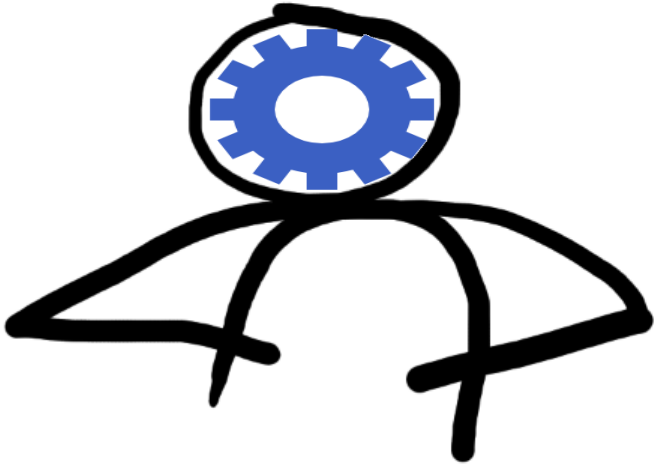


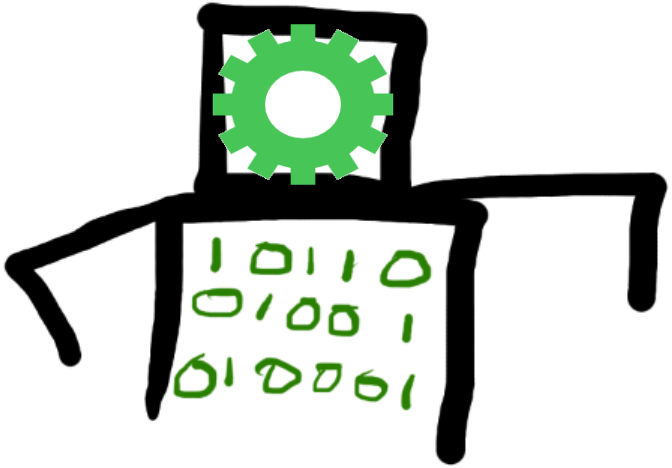
COMPUTER AS A TOOL



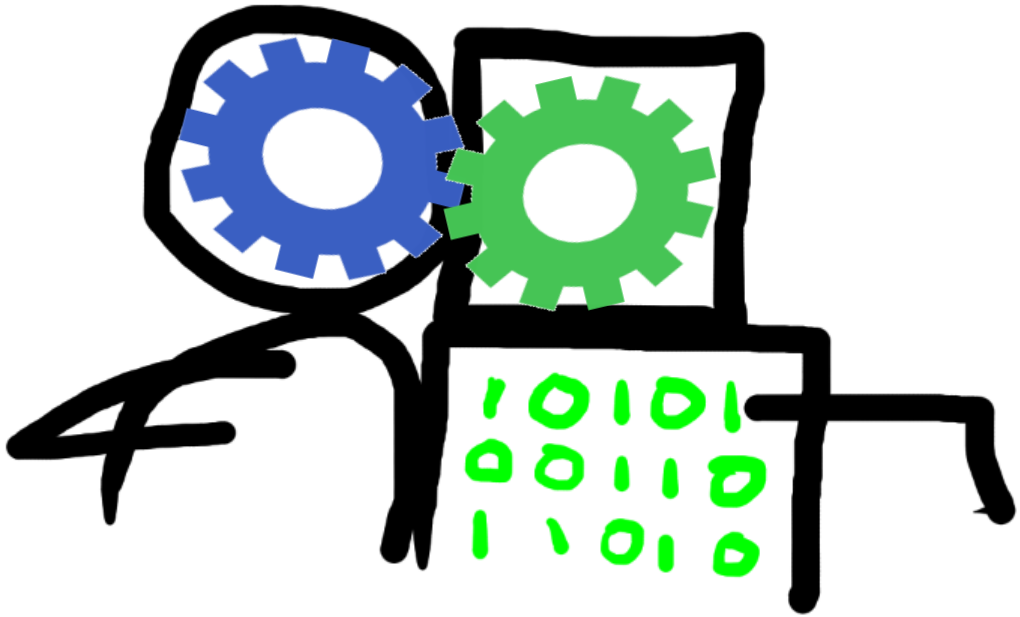


COMPUTER AS AN ISOLATED AGENT

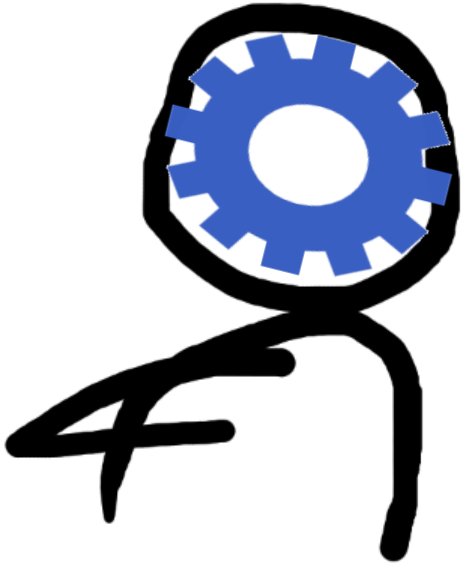


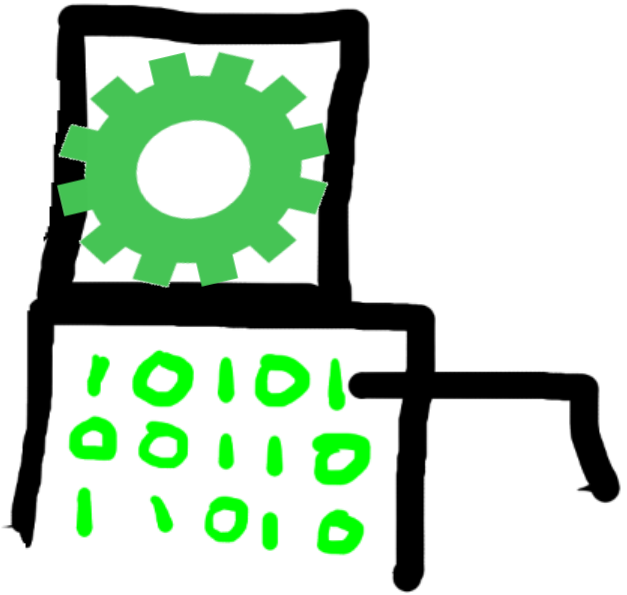


COMPUTER AS A TEAMMATE

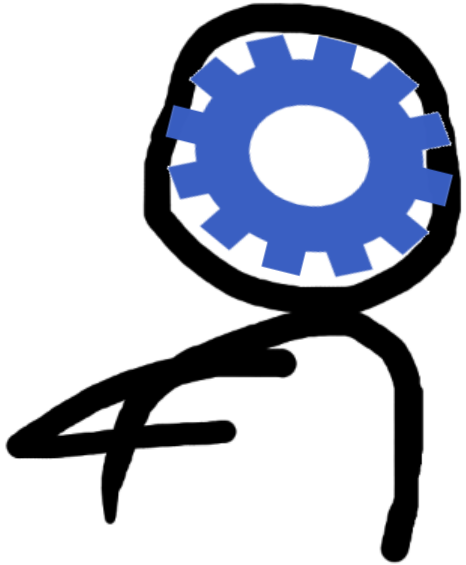


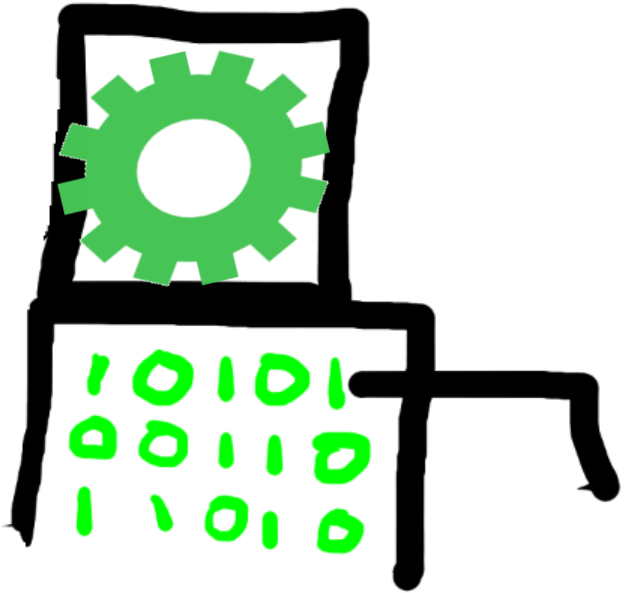
DYNAMIC TEAMING



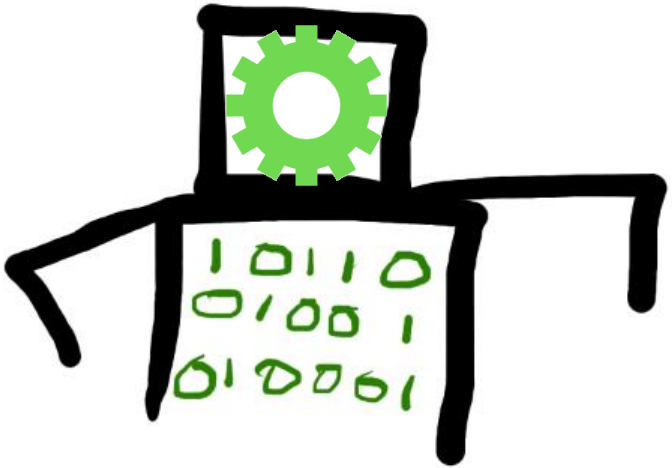


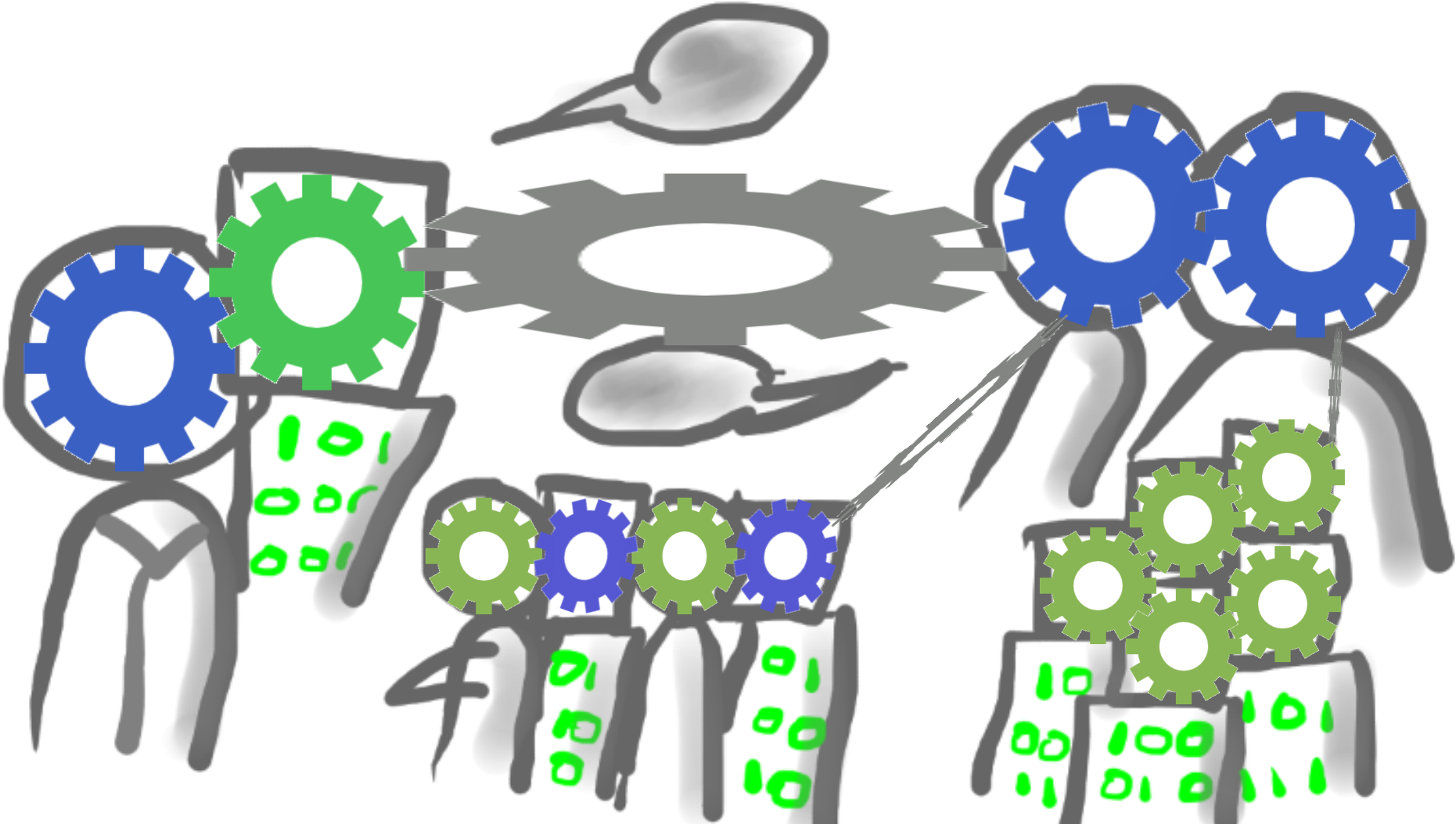
DYNAMIC TEAMING

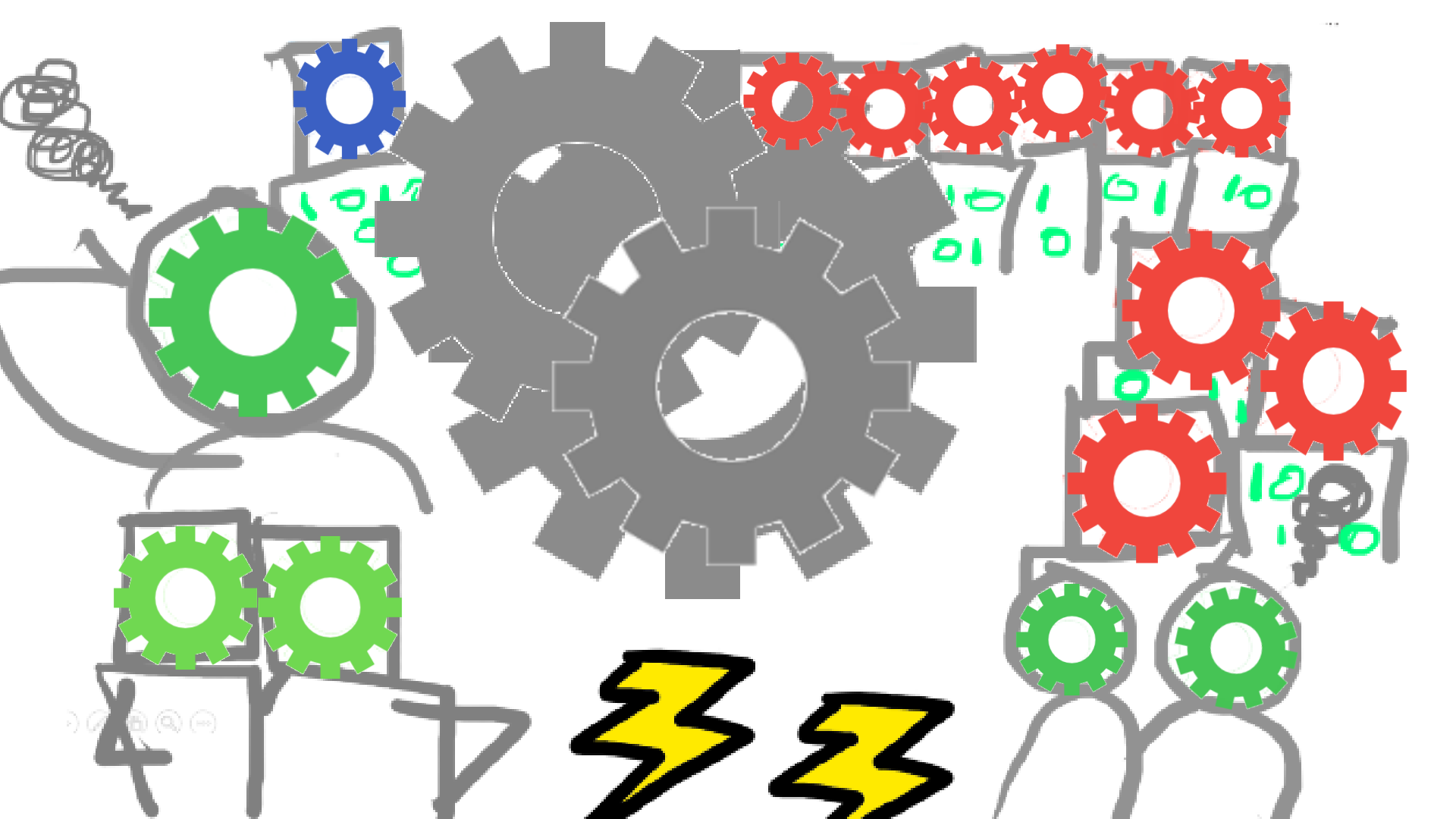




RUNAWAY AI



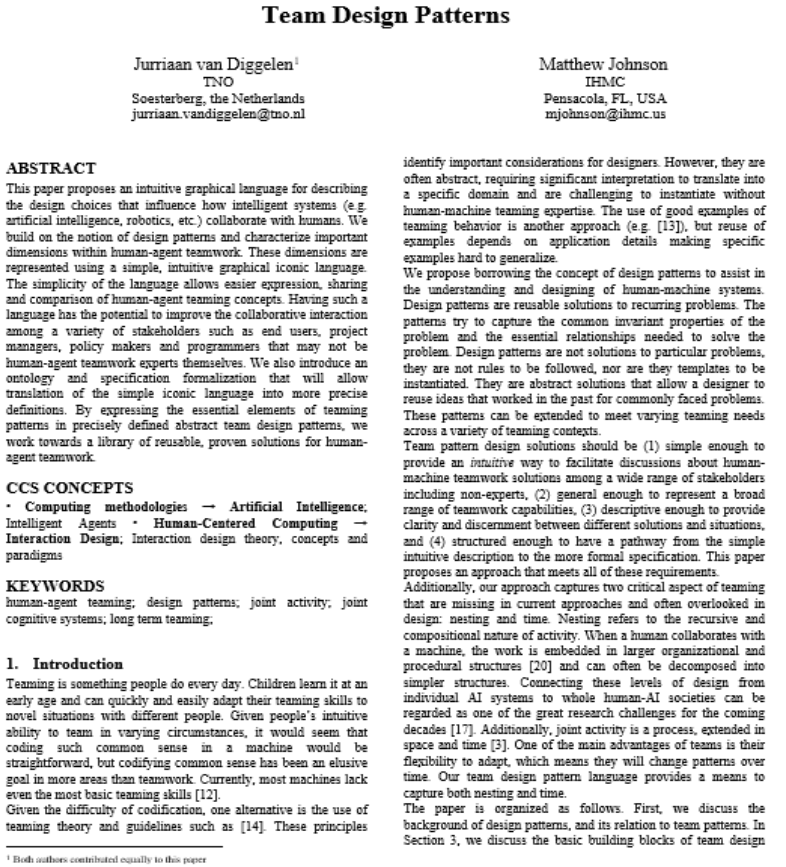




TEAM DESIGN PATTERNS

- How to design coherent human agent teams in a way that is
  - **Simple and intuitive** to allow communication among stakeholders
  - **General** enough to represent a broad range of teamwork
  - **Descriptive** enough to allow comparison of different solutions and situations
  - **Structured** enough to have a pathway from the simple intuitive description to the more formal specification.

- Fosus on:
  - **Nesting**
  - **Time**



Team Design Patterns

Jurriaan van Diggelen[1]
TNO
Soesterberg, the Netherlands
jurriaan.vandiggelen@tno.nl

Matthew Johnson
IHMC
Pensacola, FL, USA
mjohnson@ihmc.us

ABSTRACT

This paper proposes an intuitive graphical language for describing the design choices that influence how intelligent systems (e.g. artificial intelligence, robotics, etc.) collaborate with humans. We build on the notion of design patterns and characterize important dimensions within human-agent teamwork. These dimensions are represented using a simple, intuitive graphical iconic language. The simplicity of the language allows easier expression, sharing and comparison of human-agent teaming concepts. Having such a language has the potential to improve the collaborative interaction among a variety of stakeholders such as end users, project managers, policy makers and programmers that may not be human-agent teamwork experts themselves. We also introduce an ontology and specification formalization that will allow translation of the simple iconic language into more precise definitions. By expressing the essential elements of teaming patterns in precisely defined abstract team design patterns, we work towards a library of reusable, proven solutions for human-agent teamwork.

CCS CONCEPTS

• **Computing methodologies** → **Artificial Intelligence**; Intelligent Agents • **Human-Centered Computing** → **Interaction Design**; Interaction design theory, concepts and paradigms

KEYWORDS

human-agent teaming; design patterns; joint activity; joint cognitive systems; long term teaming;

1. Introduction

Teaming is something people do every day. Children learn it at an early age and can quickly and easily adapt their teaming skills to novel situations with different people. Given people's intuitive ability to team in varying circumstances, it would seem that coding such common sense in a machine would be straightforward, but codifying common sense has been an elusive goal in more areas than teamwork. Currently, most machines lack even the most basic teaming skills [12].

Given the difficulty of codification, one alternative is the use of teaming theory and guidelines such as [14]. These principles identify important considerations for designers. However, they are often abstract, requiring significant interpretation to translate into a specific domain and are challenging to instantiate without human-machine teaming expertise. The use of good examples of teaming behavior is another approach (e.g. [13]), but reuse of examples depends on application details making specific examples hard to generalize.

We propose borrowing the concept of design patterns to assist in the understanding and designing of human-machine systems. Design patterns are reusable solutions to recurring problems. The patterns try to capture the common invariant properties of the problem and the essential relationships needed to solve the problem. Design patterns are not solutions to particular problems, they are not rules to be followed, nor are they templates to be instantiated. They are abstract solutions that allow a designer to reuse ideas that worked in the past for commonly faced problems. These patterns can be extended to meet varying teaming needs across a variety of teaming contexts.

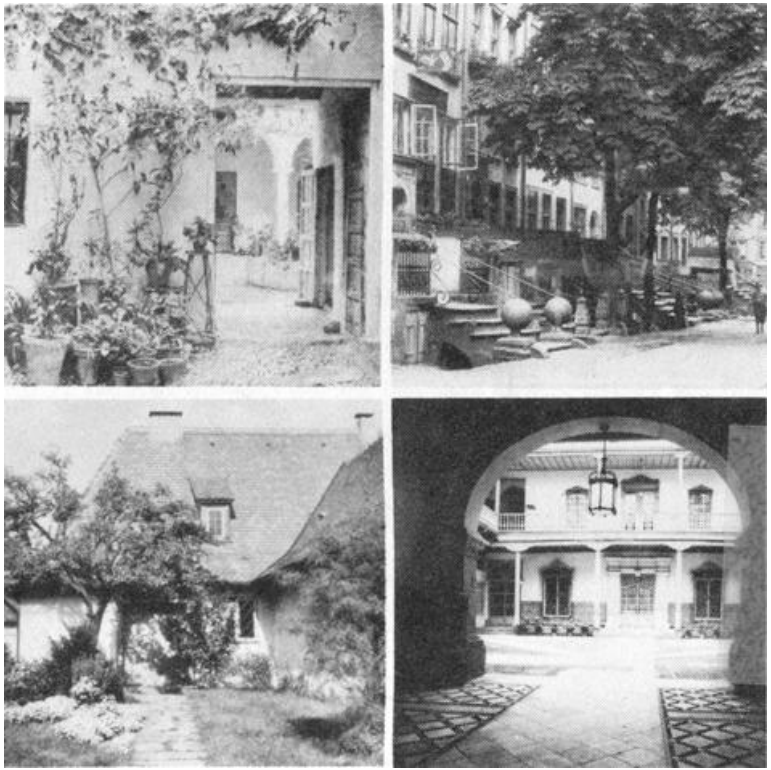
Team pattern design solutions should be (1) simple enough to provide an *intuitive* way to facilitate discussions about human-machine teamwork solutions among a wide range of stakeholders including non-experts, (2) general enough to represent a broad range of teamwork capabilities, (3) descriptive enough to provide clarity and discernment between different solutions and situations, and (4) structured enough to have a pathway from the simple intuitive description to the more formal specification. This paper proposes an approach that meets all of these requirements.

Additionally, our approach captures two critical aspect of teaming that are missing in current approaches and often overlooked in design: nesting and time. Nesting refers to the recursive and compositional nature of activity. When a human collaborates with a machine, the work is embedded in larger organizational and procedural structures [20] and can often be decomposed into simpler structures. Connecting these levels of design from individual AI systems to whole human-AI societies can be regarded as one of the great research challenges for the coming decades [17]. Additionally, joint activity is a process, extended in space and time [3]. One of the main advantages of teams is their flexibility to adapt, which means they will change patterns over time. Our team design pattern language provides a means to capture both nesting and time.

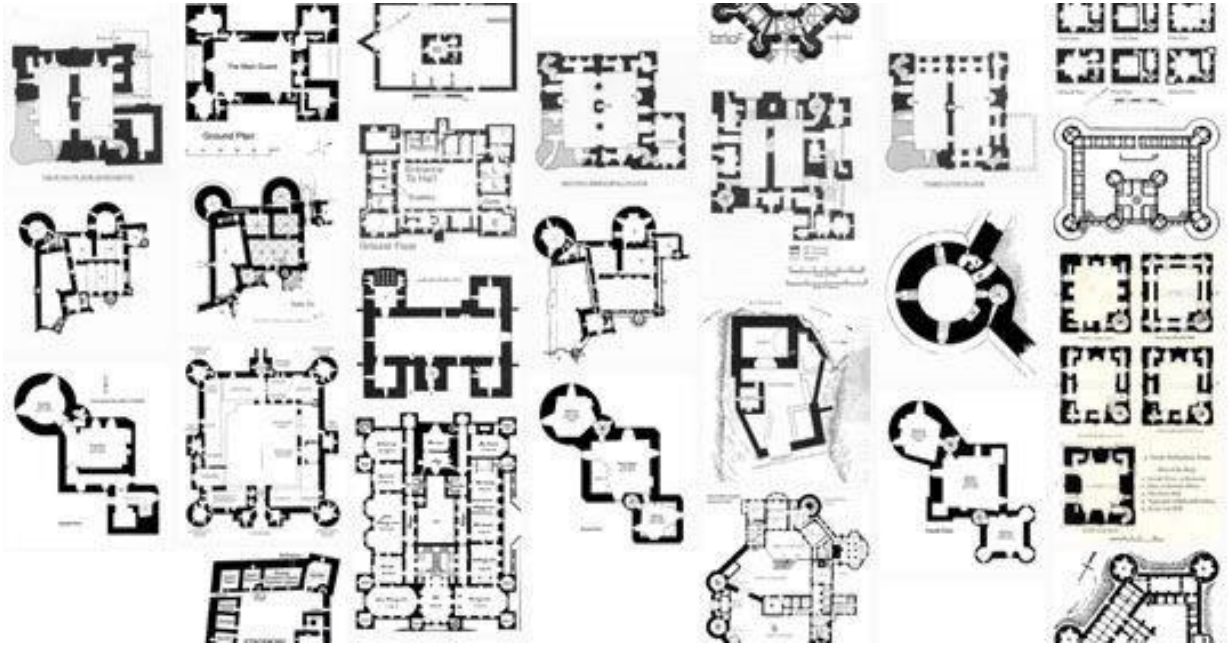
The paper is organized as follows. First, we discuss the background of design patterns, and its relation to team patterns. In Section 3, we discuss the basic building blocks of team design

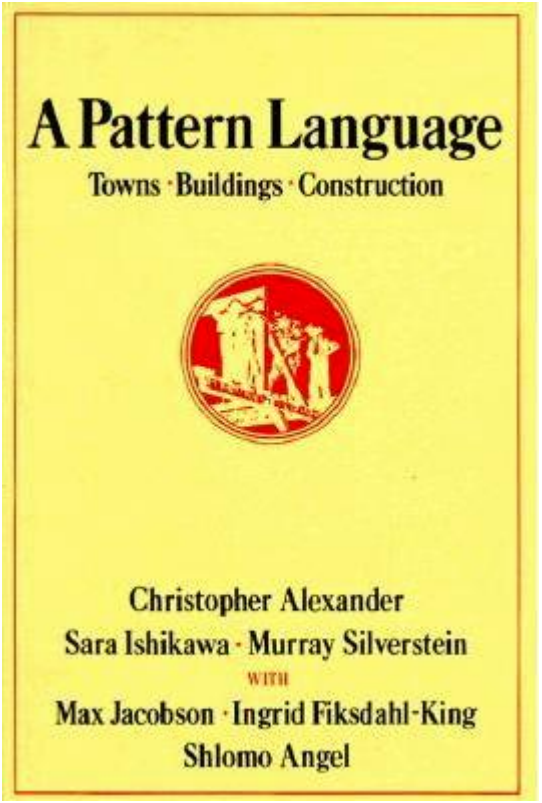
[1] Both authors contributed equally to this paper

CHRISTOPHER ALEXANDER

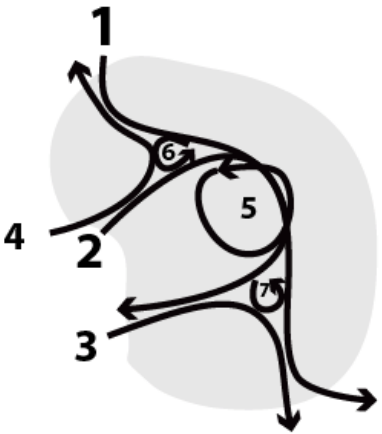


A PATTERN LANGUAGE





Christopher Alexander's Default Design Approach

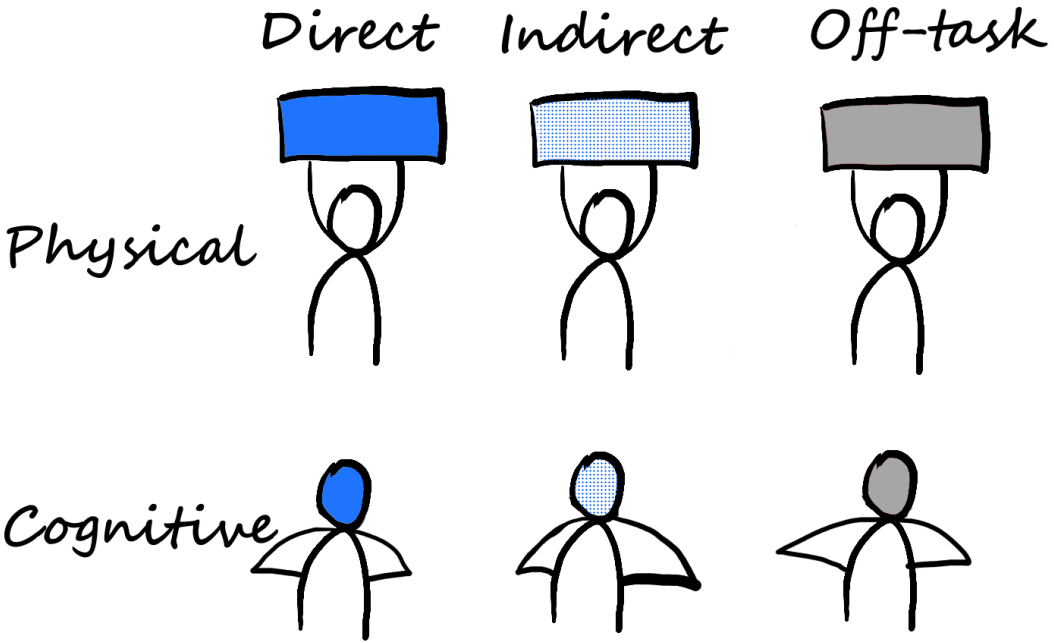


a. Start with a whole...

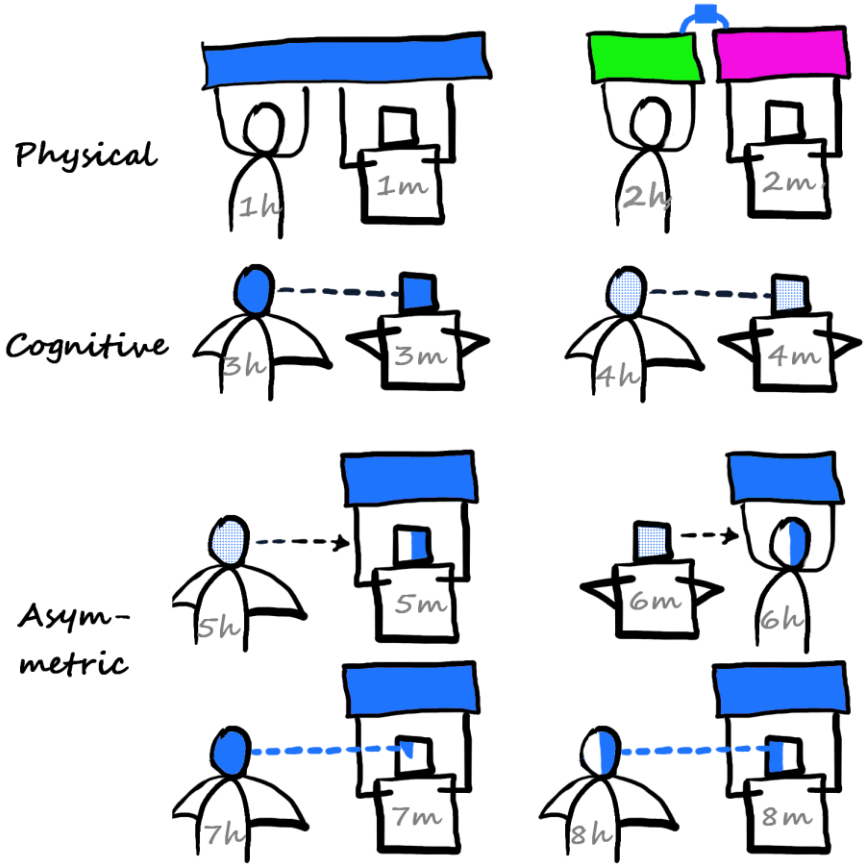
b. Differentiate it...

c. Into parts...

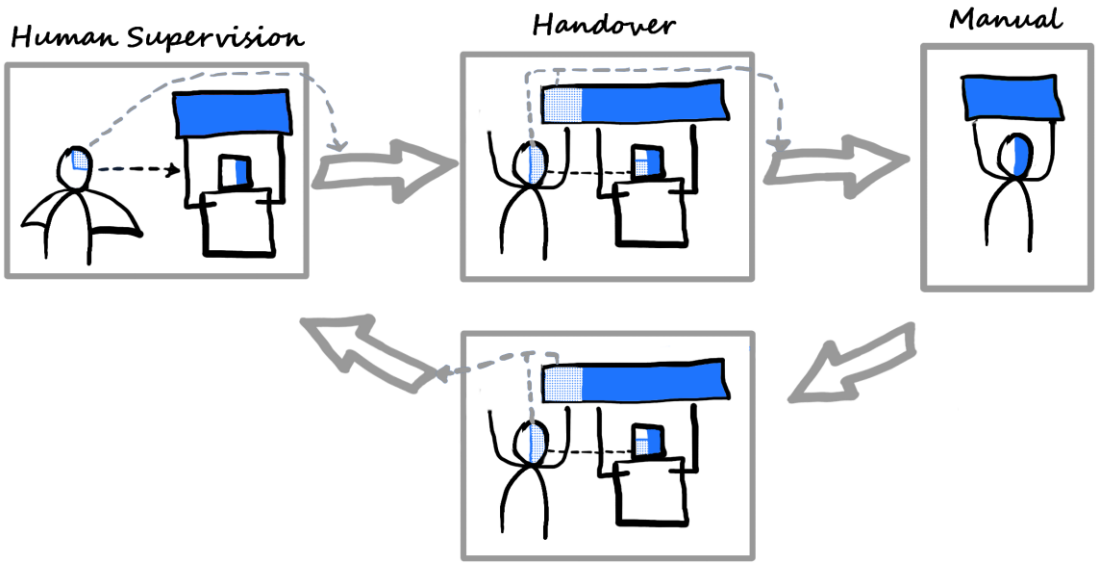
BASIC TYPES OF WORK



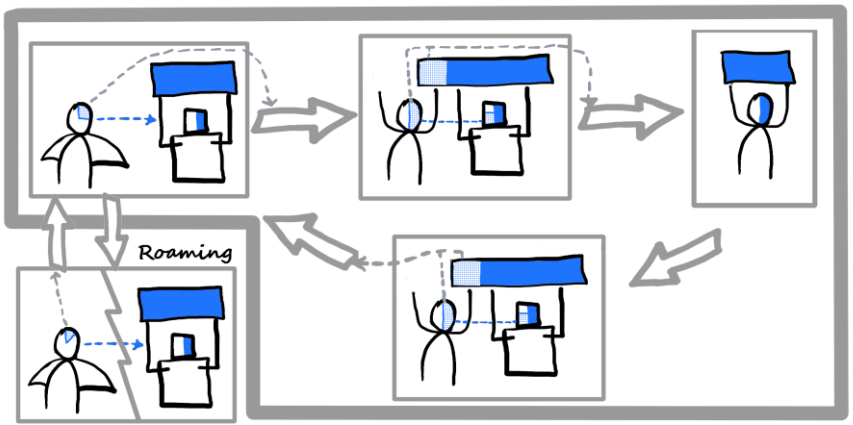
JOINT WORK

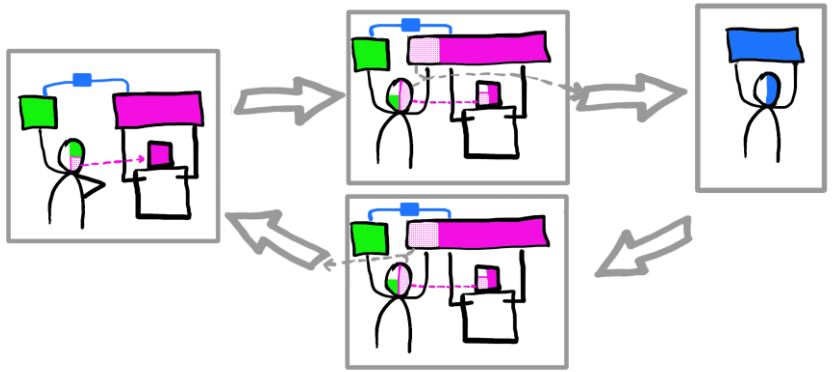

Physical
1h
1m
2h
2m
Cognitive
3h
3m
4h
4m
Asym-
metric
5h
5m
6m
6h
7h
7m
8h
8m

SUPERVISORY CONTROL

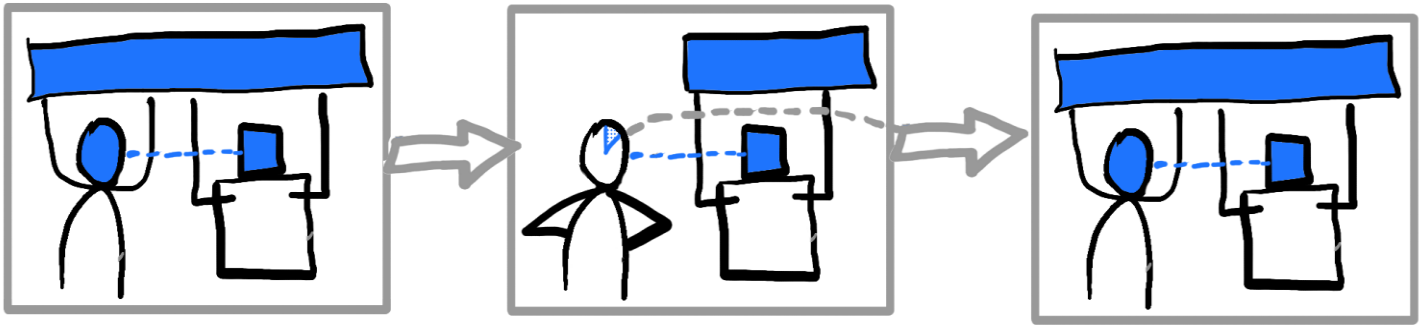


VARIANTS OF SUPERVISORY CONTROL





HIGHLY AUTONOMOUS PATTERNS



FORMAL SPECIFICATION

Teleoperation : **Team Design Pattern**

Name : "Tele operation"
Image : Img1
Use when : "machine has limited autonomous capability, and human skilled operators are available..."
Positive effect: "Clear single point of control at human operator"
Negative effect : "Imposes heavy taskload on the human"
Example : "Teleoperation of a UAV..."
Involves actors : [7h,7m]

7h : **Human**

Name: "7h"
Performs <level of engagement = 1.0> Teleoperating

Teleoperating: **Direct Cognitive Work**

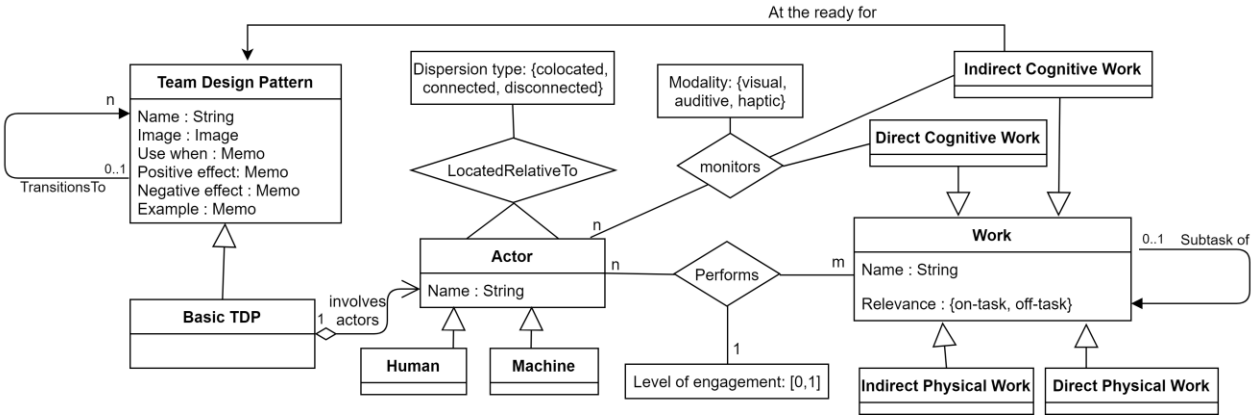
Name: "teleoperating"
Relevance = "on-task
Monitors <modality = auditive> [7h,7m]

7m : **Machine**

Name: "7m"
Performs <level of engagement = 0.1> Teleoperating
Performs <level of engagement = 1.0> PerformInstructions

PerformInstructions: **Direct physical Work**

Name: "PerformInstructions"
Relevance = "on-task



GOAL: develop a pattern library for meaningful human control.


ARTIFICIAL INTELLIGENCE
NATURAL INTELLIGENCE
QUESTIONS?
SUMANTA BARUAH