MEANINGFUL HUMAN CONTROL IN MILITARY APPLICATIONS, A NATO PERSPECTIVE

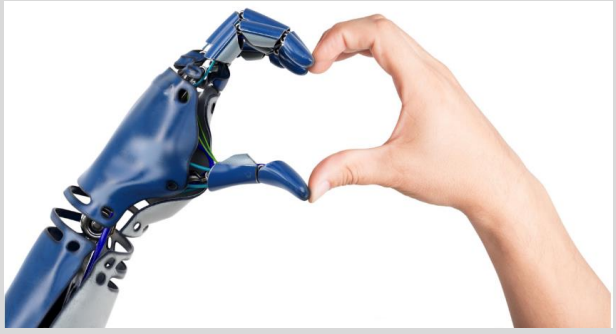
AI TECH AGORA, 26 FEBRUARY 2020

Dr. Jurriaan van Diggelen

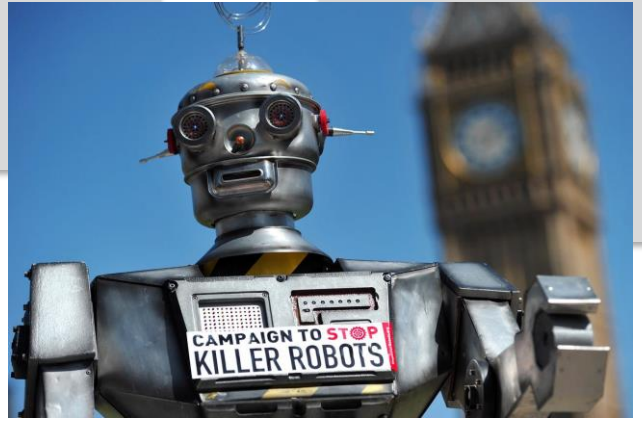
TNO innovation for life

MILITARY APPLICATIONS

USE AI FOR GOOD PURPOSES



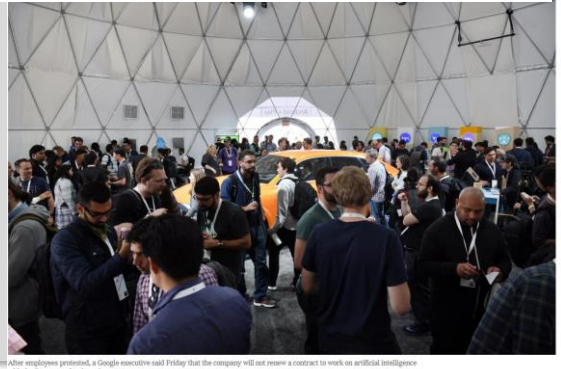
"AI should cultivate Human well-being"



The New York Times

June 1, 2018

Google Will Not Renew Pentagon Contract That Upset Employees



After employees protested, a Google executive said Friday that the company will not renew a contract to work on artificial intelligence with the Pentagon after it expires next year. Michael Short/Bloomberg

MY ACTIVITIES AT TNO

- Project leader of Dutch Defense research projects on
  - Human-machine teaming
  - Meaningful Human Control
- Chair of Nato activities on Meaningful Human Control
- Coordinating Manning and Automation portfolio for Dutch Navy

Why?

- To help Dutch Defense, Safety and Security
- To help preventing harmful and unethical applications of AI
- To help adopting multiple perspectives

DUTCH DEFENSE

Jurriaan van Diggelen (TNO) - Human-Machine Teaming

Hijacked ship's crew rescued from pirates near Somalia

9 April 2017

Share

HOW FAR DO WE WANT TO GO?

(MILITARY) AI REQUIRES MULTIPLE PERSPECTVES

- Military AI roadmaps heavily rely on dual use technologies
- Applications require diverse perspectives from
  - Operational users, Academia, Industry
  - Multi-disciplinary: AI, Systems engineering, Human Factors, Legal, Ethics…
  - Multiple countries, ethnicities, etc.

NATO PERSPECTIVE

Year founded:
1949

NATO has **2** official languages (English and French)

Number of member nations: **29**

Article **5** lays out the principle of collective defense

Number of members without an army: **1** (Iceland)

Organisation & Agencies

Organisations and agencies
- **Support**
  - NATO Support and Procurement Agency (NSPA)
    - Logistics Operations segment
    - Central Europe Pipeline System Programme
    - NATO Airlift Management Programme
- **Communications and information**
  - NATO Communications and Information Agency (NCI Agency)
    - *NATO C3 Organisation*
    - *NATO CIS Services Agency (NCSA)*
    - *NATO Consultation, Command and Control Agency (NC3A)*
- **Science and Technology (S&T)**
  - NATO Science and Technology Organization (STO)
    - *Science and Technology Board*
    - *Programme Office for Collaborative S&T*
    - Centre for Maritime Research and Experimentation (CMRE)
  - NATO Standardization Office (NSO)
  - NATO Term - The Official NATO Terminology Database
- **Programme Offices**
  - NATO Alliance Ground Surveillance Management Agency (NAGSMA)
  - NATO Helicopter Design and Development Production and Logistics Management Agency (NAHEMA)
  - NATO Medium Extended Air Defence System Design and Development, Production and Logistics Management Agency (NAMEADSMA)
  - NATO Sea Sparrow Surface Missile System Office
  - NATO Airborne Early Warning and Control Programme Management Agency (NAPMA)
- **Civil Emergency Planning**
  - Senior Civil Emergency Planning Committee (SCEPC)
  - Civil Emergency Planning Boards and Committees
  - Euro-Atlantic Disaster Response Coordination Centre (EADRCC)
- **Air Traffic Management, Air Defence**
  - Aviation Committee
  - The NATO Air Defence Committee (NADC)
  - Military Committee Air Defence Study Working Group (MC-ADSWG)
  - NATO Programming Centre (NPC)
- **Airborne Early Warning**
  - The NATO Airborne Early Warning and Control Programme Management Organisation (NAPMO)
- **Electronic Warfare**
  - NATO Electronic Warfare Advisory Committee (NEWAC)
- **Meteorology**
  - Military Committee Meteorological Group (MCMG)
- **Military Oceanography**
  - The Military Oceanography (MILOC) Group
- **Education and Training**
  - NATO Defense College (NDC)
  - The NATO School - Oberammergau, Germany
  - NATO Communications and Information Systems (NCISS) School
  - The NATO Training Group (NTG)
  - NATO Maritime Interdiction Operational Training Center - NMIOTC - Souda Naval Base, Crete, GR
- **NATO accredited Centres of Excellence**
  - Centre for Analysis and Simulation for the Preparation of Air Operations (CASPOA)
  - Civil Military Cooperation (CIMIC)
  - Cold Weather Operations (CWO)
  - Combined Joint Operations from the Sea (CJOS)
  - Command & Control (C2)
  - Cooperative Cyber Defence
  - Counter Improvised Explosive Devices (CIED)
  - Counter-Intelligence Centre of Excellence
  - Crisis Management and Disaster Response Centre of Excellence (CMDR COE)
  - Defence Against Terrorism (DAT)
  - Energy Security Centre of Excellence
  - Explosive Ordnance Disposal (EOD)
  - Human Intelligence (HUMINT)
  - Joint Air Power Competence Centre (JAPCC)
  - Joint Chemical, Biological, Radiological, & Nuclear Defence (JCBRN)
  - Military Engineering (MILENG)
  - Military Medicine (MILMED)
  - Military Police (MP)
  - Modelling and Simulation (M&S)
  - Mountain Warfare (MW)
  - Naval Mine Warfare (NMW)
  - Operations in Confined and Shallow Waters (CSW)
  - Security Force Assistance
  - Stability Policing Centre of Excellence
  - Strategic Communications Centre of Excellence (COE STRATCOM)
- **Project Steering Committees/Project Offices**
  - Alliance Ground Surveillance Capability Provisional Project Office (AGS/PPO)
  - Battlefield Information Collection and Exploitation System (BICES)
  - NATO Continuous Acquisition and Life Cycle Support Office (CALS)
  - NATO FORACS Office
  - Munitions Safety Information Analysis Center (MSIAC)

Military

Military structure
- **The Military Committee**
  - **International Military Staff**
    - Intelligence (INT)
    - Operations and Plans (OP&P)
    - Policy and Capabilities (P&C)
    - Cooperative Security (CS)
    - Logistics and Resources (L&R)
    - NATO Headquarters C3 Staff (NHQC3S)
    - NATO Situation Centre (SITCEN)
  - **Allied Command Operations (ACO)**
    - Supreme Headquarters Allied Power Europe (SHAPE) - Mons, BE
    - Headquarters Allied Joint Force Command HQ Brunssum - Brunssum, NL
      - Resolute Support
    - Headquarters Allied Joint Force Command Naples - Naples, IT
      - NATO Military Liaison Office Belgrade
      - NATO Headquarters Sarajevo (NHQSa) - Sarajevo, BA
      - NATO Headquarters Skopje (NHQSk) - Skopje, North Macedonia
    - Headquarters Allied Maritime Command HQ - Northwood, GB
      - Standing NATO Maritime Groups
        - Standing NATO Maritime Group 1 (SNMG1)
        - Standing NATO Maritime Group 2 (SNMG2)
        - Standing NATO Mine Countermeasures Group 1 (SNMCMG1)
        - Standing NATO Mine Countermeasures Group 2 (SNMCMG2)
      - MARCOM Subordinate Commands
        - NATO Maritime Air Command (COMMARAIR)
        - NATO Submarine Command (COMSUBNATO)
      - NATO Shipping Centre
    - Headquarters Allied Air Command (HQ AIRCOM) - Ramstein, DE
      - Combined Air Operations Centre (CAOC) Torrejón
      - Combined Air Operations Centre (CAOC) Uedem
      - Deployable Air Command and Control Centre (DACCC)
    - Headquarters Allied Land Command - Izmir, TR
    - The NATO CIS Group
    - Rapidly Deployable Corps Headquarters
      - Headquarters Allied Command Europe Rapid Reaction Corps (ARRC) - Innsworth-Gloucester, UK
      - Headquarters EUROCORPS in Strasbourg, FR
      - Multinational Corps Northeast - Szczecin, PL
      - Headquarters Rapid Deployable Corps Italy - Milan, IT
      - Headquarters Rapid Deployable Corps Turkey - Istanbul, TR
      - Headquarters Rapid Deployable German-Netherlands Corps - Münster, DE
      - Headquarters Rapid Deployable Corps Spain - Valencia, ES
      - Headquarters Rapid Deployable Corps France in Lille, FR
      - Headquarters Rapid Deployable Corps Greece in Thessaloniki, GR
    - Other Staffs and Commands Responsible to SACEUR
      - Immediate Reaction Forces (Maritime)
      - Naval Striking and Support Forces - STRIKFORNATO - Lisbon, PT
      - NATO Airborne Early Warning & Control Force (HQ NAEW&C Force GK)
  - **Allied Command Transformation (ACT)**
    - Headquarters Supreme Allied Commander Transformation (HQ SACT) - Norfolk, US
    - joint Warfare Centre (JWC) - Stavanger, NO
    - NATO Training Group
    - NATO Joint Force Training Centre (JFTC) - Bydgoszcz, PL
    - NATO Maritime Interdiction Operational Training Center (NMIOTC) - Souda Naval Base, Crete, GR
    - NATO School - Oberammergau, DE
    - Joint Analysis and Lessons Learned Centre - Lisbon, PT
  - **Other NATO Command & Staff Organisations**
    - Canada-US Regional Planning Group (CUSRPG)
    - Combined Joint Planning Staff (CJPS) - Mons, BE

Civilian

Civilian structure
- NATO Headquarters
- Permanent Representatives and National Delegations
- International Staff (IS)
  - **Private Office (PO)**
    - Special Representative on Women, Peace and Security
  - **Political Affairs and Security Policy Division**
    - NATO Liaison Office (NLO) in Georgia
    - Office of the NATO Liaison Officer (NLO) in Central Asia
    - Partnership for Peace Documentation Center
  - **Operations Division**
    - Civil Emergency Planning
    - Euro-Atlantic Disaster Response Coordination Centre (EADRCC)
    - NATO Situation Centre
    - NATO Advisory and Liaison Team (NALT)
  - **Emerging Security Challenges Division**
    - Science for Peace and Security
    - The Weapons of Mass Destruction Non-Proliferation Centre (WMDC)
    - Defence Against Terrorism Programme of Work (DAT POW)
  - **Defence Policy and Planning Division**
    - NATO Liaison Office (NLO) Ukraine
  - **Defence Investment Division**
    - The Conference of National Armaments Directors (AC/259)
      - Group of National Directors on Codification (AC/135)
      - NATO Naval Armaments Group (NNAG) (AC/141)
      - NATO Air Force Armaments Group (NAFAG) (AC/224)
        - The Joint Capability Group Intelligence, Surveillance and Reconnaissance (JCGISR)
      - NATO Army Armaments Group (NAAG) (AC/225)
      - Alliance Future Surveillance and Control Project Group (AFSC PG)
      - Ballistic Missile Defence (BMD)
      - CNAD Ammunition Safety Group (AC/326)
      - Life Cycle Management Group (AC/327)
      - NATO Industrial Advisory Group (NIAG)
      - C-IED Activities
      - Industry Relations
    - NATO Naval Forces Sensor and Weapons Accuracy Check Sites (FORACS)
    - Air and Missile Defence Committee (AMDC) (AC/336)
    - Aviation Committee (AVC) (AC/92) *site is under construction*
    - Munitions Safety Information Analysis Center (MSIAC)
  - **Public Diplomacy Division**
    - Co-sponsorship grants
    - NATO Information and Documentation Centre, Kyiv, Ukraine
    - NATO Information Office in Moscow (in Russian)
    - NATO Contact Point Embassies in partner countries
  - **Executive Management**
    - Archives
    - NATO Internship Programme
    - NATO Multimedia Library
    - Recruitment Service
  - **Joint Intelligence and Security Division (JIS)**
  - **NATO Office of Resources (NOR)**
  - **NATO Office of Security (NOS)**
  - NATO Headquarters Consultation, Command and Control Staff (NHQC3S)
  - Office of the Financial Controller (FinCon)
  - Office of the Chairman of the Senior Resource Board (SRB)
  - Office of the Chairman of the Civil and Military Budget Committees (CBC/MBC))
  - International Board of Auditors for NATO (IBAN)
  - NATO Administrative Tribunal
  - NATO Production and Logistics Organisations (NPLO)

Organisation & Agencies

Organisations and agencies
- **Support**
  - NATO Support and Procurement Agency (NSPA)
    - Logistics Operations segment
    - Central Europe Pipeline System Programme
    - NATO Airlift Management Programme
- **Communications and information**
  - NATO Communications and Information Agency (NCI Agency)
    - *NATO C3 Organisation*
    - *NATO CIS Services Agency (NCSA)*
    - *NATO Consultation, Command and Control Agency (NC3A)*
- **Science and Technology (S&T)**
  - NATO Science and Technology Organization (STO)
    - *Science and Technology Board*
    - *Programme Office for Collaborative S&T*
    - Centre for Maritime Research and Experimentation (CMRE)
  - NATO Standardization Office (NSO)
  - NATO Term - The Official NATO Terminology Database
- **Programme Offices**
  - NATO Alliance Ground Surveillance Management Agency (NAGSMA)
  - NATO Helicopter Design and Development Production and Logistics Manageme
  - NATO Medium Extended Air Defence System Design and Development, Produc
  - NATO Sea Sparrow Surface Missile System Office
  - NATO Airborne Early Warning and Control Programme Management Agency (N
- **Civil Emergency Planning**
  - Senior Civil Emergency Planning Committee (SCEPC)
  - Civil Emergency Planning Boards and Committees
  - Euro-Atlantic Disaster Response Coordination Centre (EADRCC)
- **Air Traffic Management, Air Defence**
  - Aviation Committee
  - The NATO Air Defence Committee (NADC)
  - Military Committee Air Defence Study Working Group (MC ADSWG)
  - NATO Programming Centre (NPC)
- **Airborne Early Warning**
  - The NATO Airborne Early Warning and Control Programme Management Organisation (NAPMO)
- **Electronic Warfare**
  - NATO Electronic Warfare Advisory Committee (NEWAC)
- **Meteorology**
  - Military Committee Meteorological Group (MCMG)
- **Military Oceanography**
  - The Military Oceanography (MILOC) Group
- **Education and Training**
  - NATO Defense College (NDC)
  - The NATO School - Oberammergau, Germany
  - NATO Communications and Information Systems (NCISS) School
  - The NATO Training Group (NTG)
  - NATO Maritime Interdiction Operational Training Center - NMIOTC - Souda Naval Base, Crete, GR
- **NATO accredited Centres of Excellence**
  - Centre for Analysis and Simulation for the Preparation of Air Operations (CASPOA)
  - Civil Military Cooperation (CIMIC)
  - Cold Weather Operations (CWO)
  - Combined Joint Operations from the Sea (CJOS)
  - Command & Control (C2)
  - Cooperative Cyber Defence
  - Counter Improvised Explosive Devices (CIED)
  - Counter-Intelligence Centre of Excellence
  - Crisis Management and Disaster Response Centre of Excellence (CMDR COE)
  - Defence Against Terrorism (DAT)
  - Energy Security Centre of Excellence
  - Explosive Ordnance Disposal (EOD)
  - Human Intelligence (HUMINT)
  - Joint Air Power Competence Centre (JAPCC)
  - Joint Chemical, Biological, Radiological, & Nuclear Defence (JCBRN)
  - Military Engineering (MILENG)
  - Military Medicine (MILMED)
  - Military Police (MP)
  - Modelling and Simulation (M&S)
  - Mountain Warfare (MW)
  - Naval Mine Warfare (NMW)
  - Operations in Confined and Shallow Waters (CSW)
  - Security Force Assistance
  - Stability Policing Centre of Excellence
  - Strategic Communications Centre of Excellence (COE STRATCOM)
- **Project Steering Committees/Project Offices**
  - Alliance Ground Surveillance Capability Provisional Project Office (AGS/PPO)
  - Battlefield Information Collection and Exploitation System (BICES)
  - NATO Continuous Acquisition and Life Cycle Support Office (CALS)
  - NATO FORACS Office
  - Munitions Safety Information Analysis Center (MSIAC)

NATO Science and Technology Organisation

Centres of Excellence

Military

Military structure
- **The Military Committee**
  - **International Military Staff**
    - Intelligence (INT)
    - Operations and Plans (OP&P)
    - Policy and Capabilities (P&C)
    - Cooperative Security (CS)
    - Logistics and Resources (L&R)
    - NATO Headquarters C3 Staff (NHQC3S)
    - NATO Situation Centre (SITCEN)
  - **Allied Command Operations (ACO)**
    - Supreme Headquarters Allied Power Europe (SHAPE) - Mons, BE
    - Headquarters Allied Joint Force Command HQ Brunssum - Brunssum, NL
      - Resolute Support
    - Headquarters Allied Joint Force Command Naples - Naples, IT
      - NATO Military Liaison Office Belgrade
      - NATO Headquarters Sarajevo (NHQSa) - Sarajevo, BA
      - NATO Headquarters Skopje (NHQSk) - Skopje, North Macedonia
    - Headquarters Allied Maritime Command HQ - Northwood, GB
      - Standing NATO Maritime Groups
        - Standing NATO Maritime Group 1 (SNMG1)
        - Standing NATO Maritime Group 2 (SNMG2)
        - Standing NATO Mine Countermeasures Group 1 (SNMCMG1)
        - Standing NATO Mine Countermeasures Group 2 (SNMCMG2)
      - MARCOM Subordinate Commands
        - NATO Maritime Air Command (COMMARAIR)
        - NATO Submarine Command (COMSUBNATO)
      - NATO Shipping Centre
    - Headquarters Allied Air Command (HQ AIRCOM) - Ramstein, DE
      - Combined Air Operations Centre (CAOC) Torrejón
      - Combined Air Operations Centre (CAOC) Uedem
      - Deployable Air Command and Control Centre (DACCC)
    - Headquarters Allied Land Command - Izmir, TR
    - The NATO CIS Group
    - Rapidly Deployable Corps Headquarters
      - Headquarters Allied Command Europe Rapid Reaction Corps (ARRC) - Innsworth-Gloucester, UK
      - Headquarters EUROCORPS in Strasbourg, FR
      - Multinational Corps Northeast - Szczecin, PL
      - Headquarters Rapid Deployable Corps Italy - Milan, IT
      - Headquarters Rapid Deployable Corps Turkey - Istanbul, TR
      - Headquarters Rapid Deployable German-Netherlands Corps - Münster, DE
      - Headquarters Rapid Deployable Corps Spain - Valencia, ES
      - Headquarters Rapid Deployable Corps France
      - Headquarters Rapid Deployable Corps Greece
    - Other Staffs and Commands Responsible to SACEUR
      - Immediate Reaction Forces (Maritime)
      - Naval Striking and Support Forces - STRIKFORNA
      - NATO Airborne Early Warning & Control Force
  - **Allied Command Transformation (ACT)**
    - Headquarters Supreme Allied Commander Transformation
    - joint Warfare Centre (JWC) - Stavanger, NO
    - NATO Training Group
    - NATO Joint Force Training Centre (JFTC) - Bydgoszcz, PL
    - NATO Maritime Interdiction Operational Training Center
    - NATO School - Oberammergau, DE
    - Joint Analysis and Lessons Learned Centre - Lisbon, PT
  - **Other NATO Command & Staff Organisations**
    - Canada-US Regional Planning Group (CUSRPG)
    - Combined Joint Planning Staff (CJPS) - Mons, BE

ACT - JALLC

Civilian

Civilian structure
- NATO Headquarters
- Permanent Representatives and National Delegations
- International Staff (IS)
  - **Private Office (PO)**
    - Special Representative on Women, Peace and Security
  - **Political Affairs and Security Policy Division**
    - NATO Liaison Office (NLO) in Georgia
    - Office of the NATO Liaison Officer (NLO) in Central Asia
    - Partnership for Peace Documentation Center
  - **Operations Division**
    - Civil Emergency Planning
    - Euro-Atlantic Disaster Response Coordination Centre (EADRCC)
    - NATO Situation Centre
    - NATO Advisory and Liaison Team (NALT)
  - **Emerging Security Challenges Division**
    - Science for Peace and Security
    - The Weapons of Mass Destruction Non-Proliferation Centre (WMDC)
    - Defence Against Terrorism Programme of Work (DAT POW)
  - **Defence Policy and Planning Division**
    - NATO Liaison Office (NLO) Ukraine
  - **Defence Investment Division**
    - The Conference of National Armaments Directors (AC/259)
    - Group of National Directors on Codification (AC/135)
    - NATO Naval Armaments Group (NNAG) (AC/141)
    - Air Force Armaments Group (NAFAG) (AC/224)
    - Joint Capability Group Intelligence, Surveillance and Reconnaissance (JCGISR)
    - NATO Army Armaments Group (NAAG) (AC/225)
    - Alliance Ground Surveillance and Control Project Group (AFSC PG)
    - Ballistic Missile Defence (BMD)
    - CNAD Ammunition Safety Group (AC/326)
    - Life Cycle Management Group (AC/327)
    - NATO Industrial Advisory Group (NIAG)
    - C-IED Activities
    - Industry Relations
    - NATO Naval Forces Sensor and Weapons Accuracy Check Sites (FORACS)
    - Air and Missile Defence Committee (AMDC) (AC/336)
    - Aviation Committee (AVC) (AC/92) *site is under construction*
    - Munitions Safety Information Analysis Center (MSIAC)
  - **Public Diplomacy Division**
    - Co-sponsorship grants
    - NATO Information and Documentation Centre, Kyiv, Ukraine
    - NATO Information Office in Moscow (in Russian)
    - NATO Contact Point Embassies in partner countries
  - **Executive Management**
    - Archives
    - NATO Internship Programme
    - NATO Multimedia Library
    - Recruitment Service
  - **Joint Intelligence and Security Division (JIS)**
  - **NATO Office of Resources (NOR)**
  - **NATO Office of Security (NOS)**
  - NATO Headquarters Consultation, Command and Control Staff (NHQC3S)
  - Office of the Financial Controller (FinCon)
  - Office of the Chairman of the Senior Resource Board (SRB)
  - Office of the Chairman of the Civil and Military Budget Committees (CBC/MBC))
  - International Board of Auditors for NATO (IBAN)
  - NATO Administrative Tribunal
  - NATO Production and Logistics Organisations (NPLO)

NATO Industry Advisory Group

MEANINGFUL HUMAN CONTROL

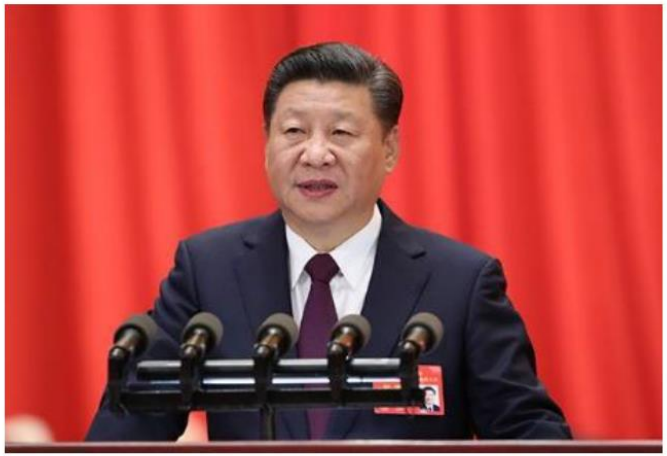
US & WORLD | TECH | ARTIFICIAL INTELLIGENCE

Putin says the nation that leads in AI 'will be the ruler of the world'

43

The Russian president warned that artificial intelligence offers 'colossal opportunities' as well as dangers

By James Vincent | @jjvincent | Sep 4, 2017, 4:53am EDT



"… by 2030, China's AI theories, technologies, and applications should achieve world leading levels, making China the world's primary AI innovation center…"

6 VOORALSNOG LIJKT CHINA DE RACE TE WINNEN

Deze mondiale race om de beste en meeste AI speelt zich vooral af tussen de VS en China, waar China *en route* is om gaan winnen. Dit blijkt uit een vergelijking van Nederland en/of Europa met China en de VS op drie vlakken.

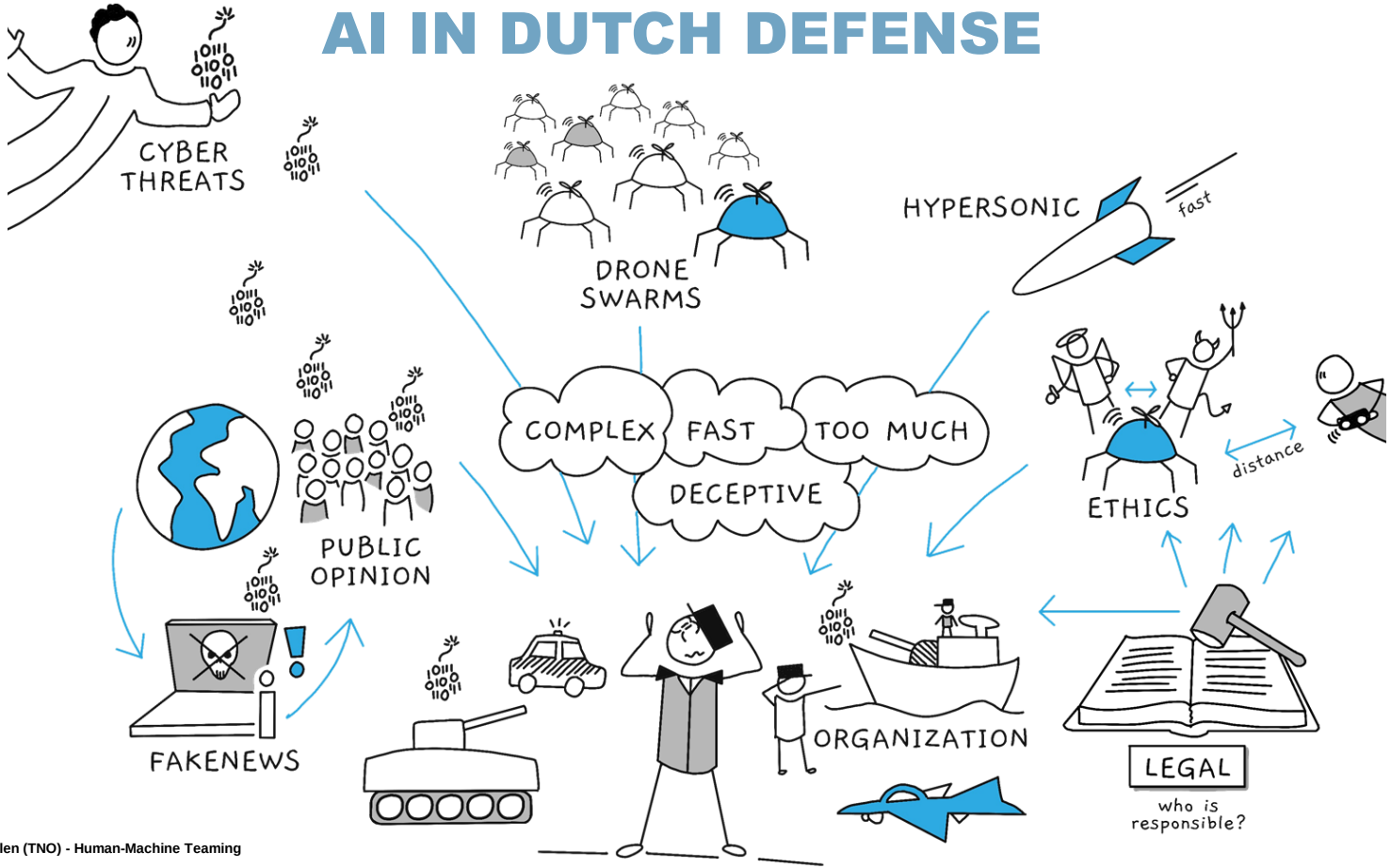
Het onderwijs in China is weliswaar kwalitatief wat minder goed dan dat van de VS en Europa, maar China compenseert dat op twee manieren. Ten eerste heeft China een relatief hoog aantal studenten in dit vakgebied. Op haar beste drie universi-

Translated from Chinese AI Development Plan, July 2017

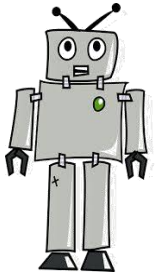
"And we believe quite strongly that the technological sauce of the Third Offset is going to be advances in Artificial Intelligence (AI) and autonomy."

Former U.S. Deputy Secretary of Defense Bob Work, April 2016

AI IN DUTCH DEFENSE



MEANINGFUL HUMAN CONTROL



The battle is too fast to allow human involvement, so we need autonomous AI!
Artificial ethics can make war more humane.

MEANINGFUL HUMAN CONTROL

Prohibit all LAWS!

Article 36

Humans should exercise control over individual attacks, not simply overall operations. Only by prohibiting the use of fully autonomous weapons can such control be guaranteed.

NATO HFM-ET-178: MHC OVER AI BASED SYSTEMS

AFRL
TNO
DSTL
NASA
Fraunhofer
FOI

WHAT IS MEANINGFUL HUMAN CONTROL?

Humans have the ability to make informed choices in sufficient time to influence AI-based systems in order to enable a desired effect or to prevent an undesired immediate or future effect on the environment.

Characteristics:

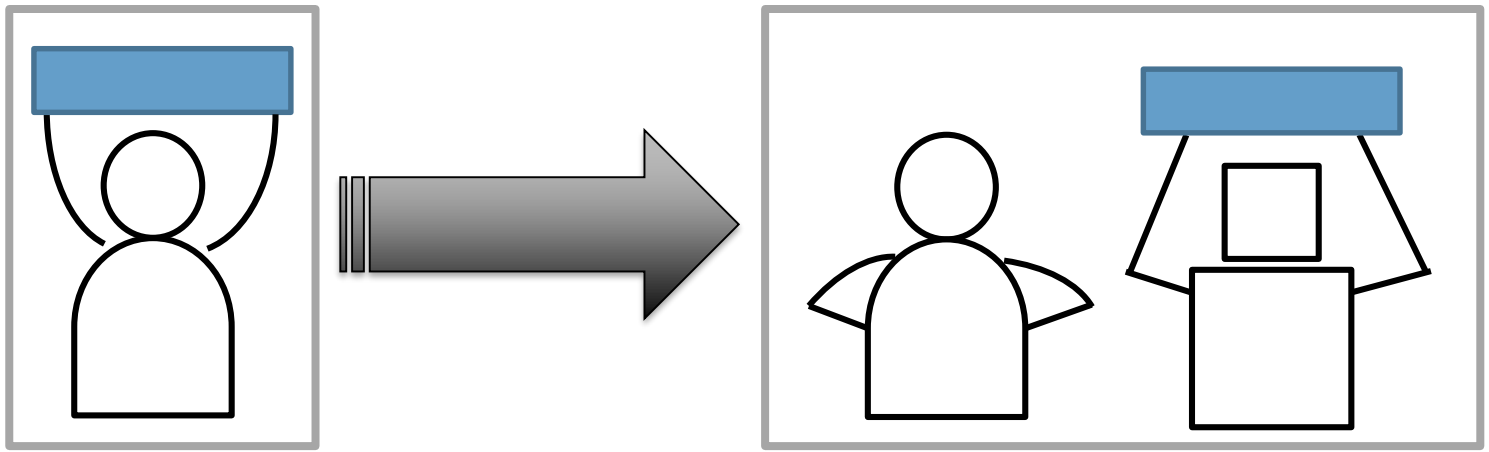
- Human has freedom of choice
- Human has ability to impact the behaviour of the system
- Human has time to decide to engage and sufficient situation, and system understanding
- Human is capable to predict the behavior of the system and the effects of the environment (physical and information)
- Influence over AI-based systems can be achieved in various ways, such as policy-making, training, HMI design, organizational design, operations, etc.
- The above encompasses cases from instantaneous (e.g. number of seconds) to very delayed response (several hours to days, e.g. before-the-loop) to control.

for meaningful human control

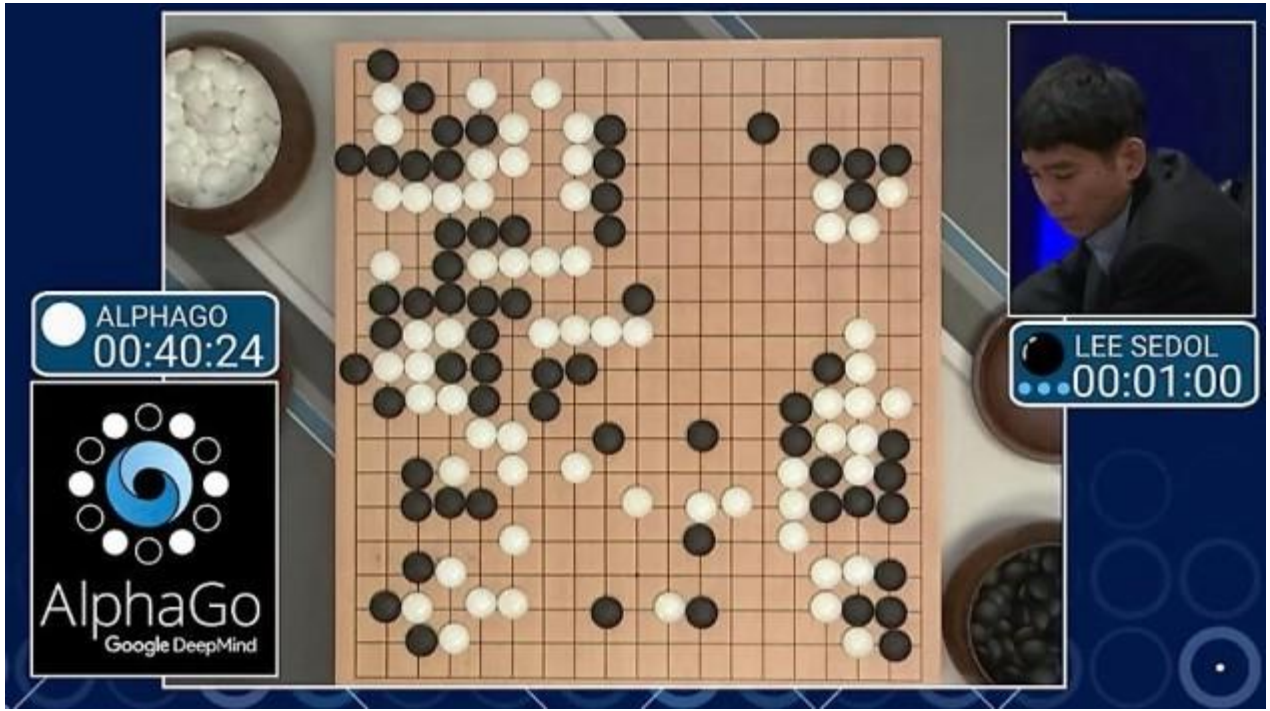
FIVE ROLES FOR AI

- AI as a substitution for human
- AI as an obedient servant
- AI as a human enhancer
- AI as work creator
- AI as a team partner

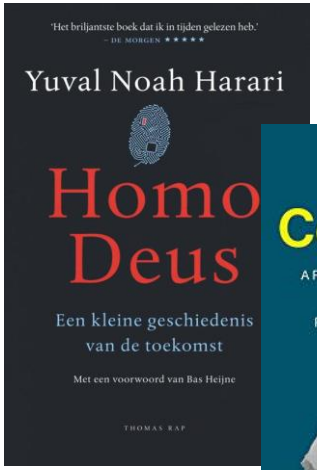
ROLE 1 AI AS A SUBSTITUTION FOR HUMANS

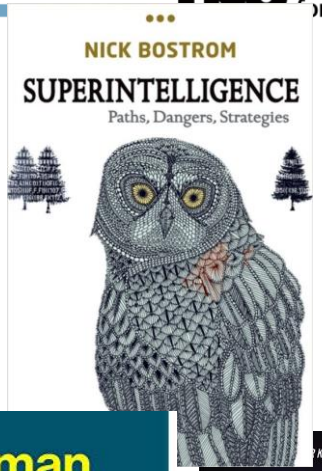


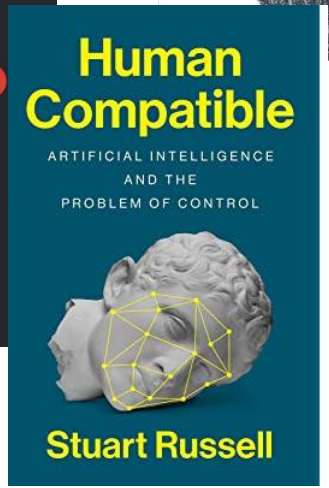
ALPHA (GO) ZERO

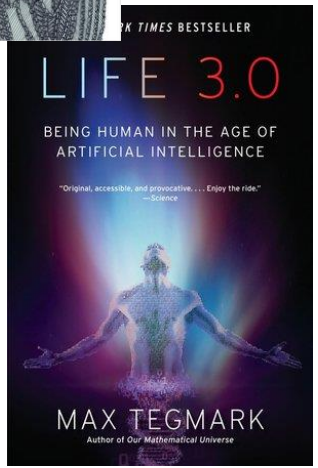


AI AS A SUBSTITUTION FOR HUMANS REQUIRES HUMAN-LEVEL INTELLIGENCE OR ABOVE

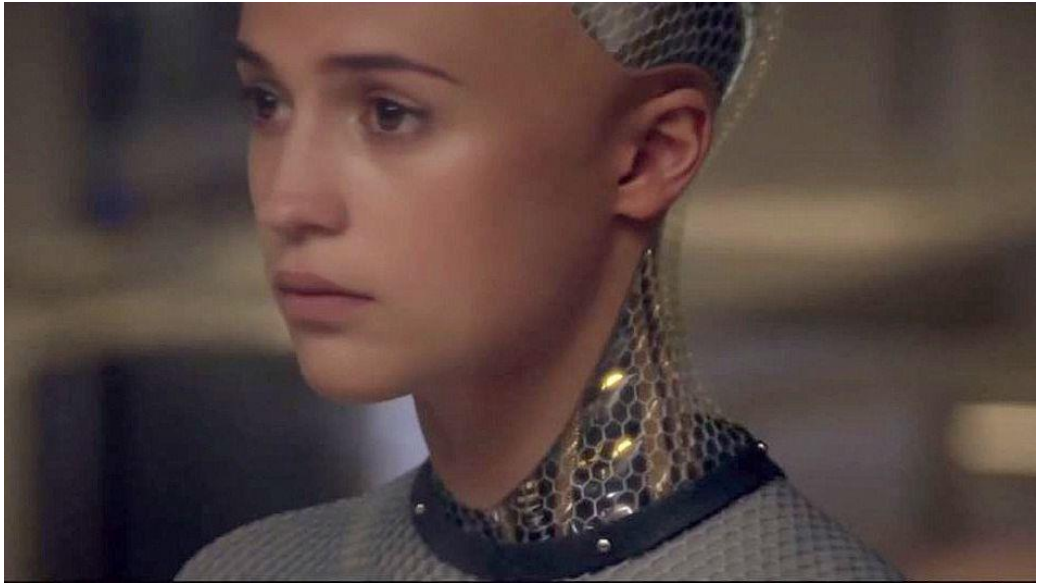






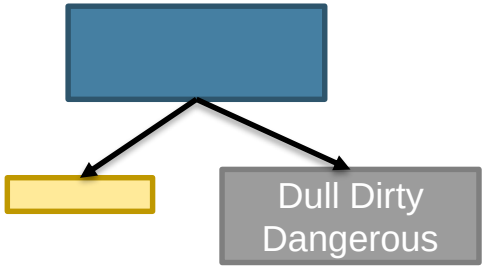


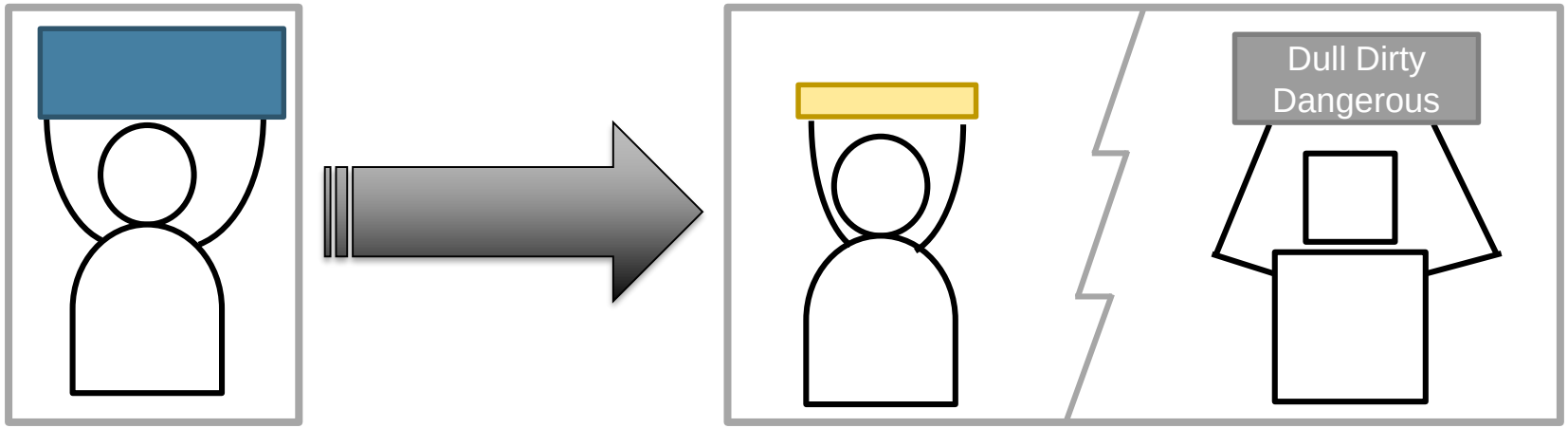
MEANINGFUL HUMAN CONTROL OF AGI



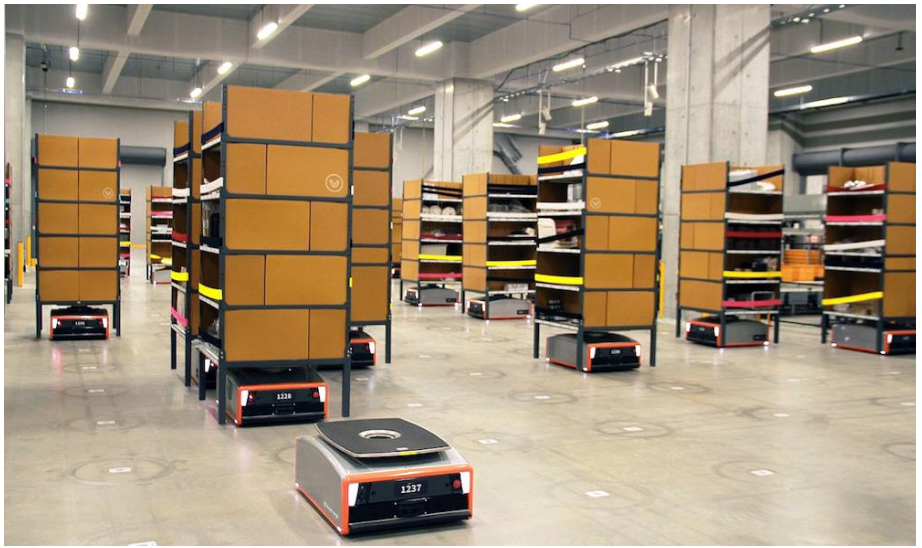
Is all about designing the proper objective function…

ROLE 2 AI AS AN OBEDIENT SERVANT





OBEDIENT SERVANT REQUIRES A PREDICTABLE ENVIRONMENT

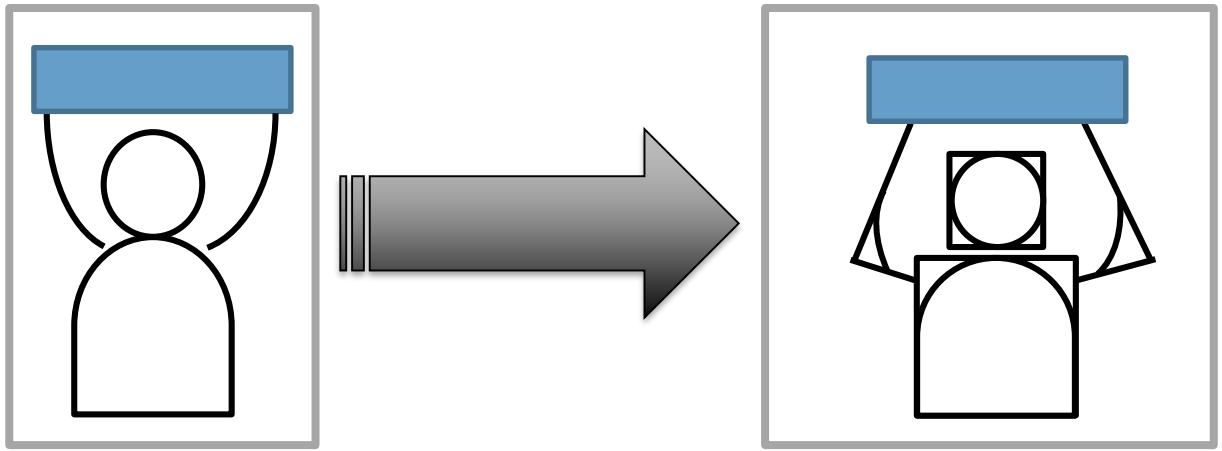


MEANINGFUL HUMAN CONTROL OF AN OBEDIENT SERVANT

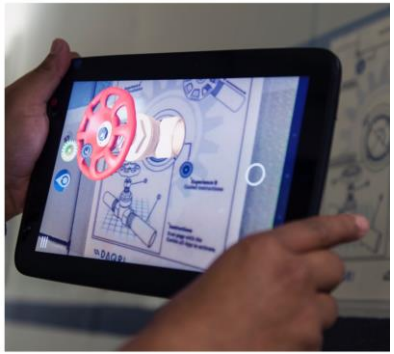
Samsung SGR-A1

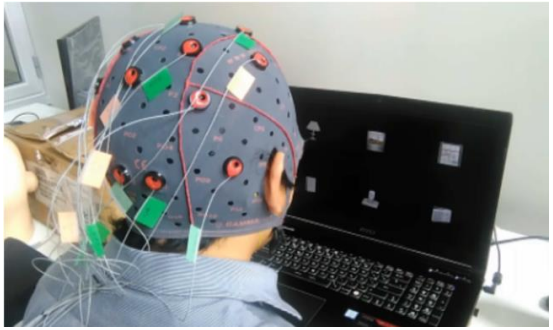
- **Laser rangefinder**
- **IR thermographic camera**
- **Mounted weapons (stock)**
  - 5.56 x 45-mm automatic light machine-gun
  - A lightweight 40-mm multiple-grenade launcher.

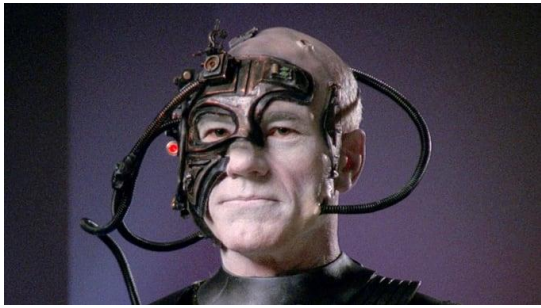
ROLE 3: AI AS A HUMAN ENHANCER

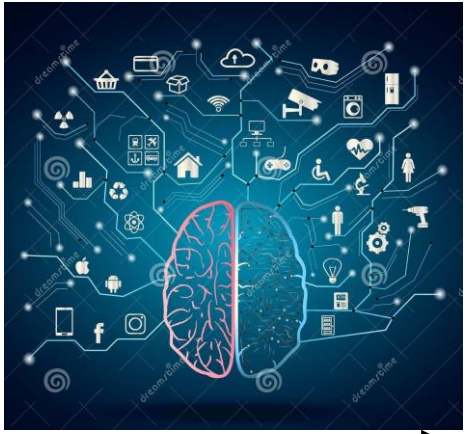


HUMAN MACHINE INTEGRATION









Now

Scifi

AI AS HUMAN ENHANCER IS STILL IN ITS INFANCY

Annu. Rev. Biophys. Bioeng. 1973.2:157-180. Downloaded from www.annualreviews.org
Access provided by TNO Human Factors - LIBRARY INSTITUTE FOR PERCEPTION TNO on 02/14/20. For personal use only.

Copyright 1973. All rights reserved

TOWARD DIRECT BRAIN-COMPUTER COMMUNICATION 9027

JACQUES J. VIDAL[1]
Brain Research Institute, University of California, Los Angeles, California

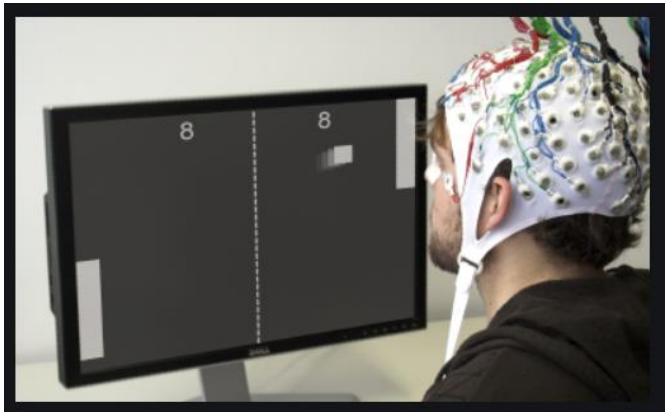
Electroencephalographic or EEG signals collected on the human scalp are sustained fluctuations of electrical potential that reflect corresponding variations in the upper layers of the brain cortex below the scalp surface. The signal structure is that of a stochastic time series with almost stationary epochs of various lengths separated by sharper transitions or disruptions. Amplitudes are small (up to a few tens of microvolts) and spectral decomposition reveals that very little power remains at frequencies above 30 Hz. Most of it is contained at very low frequencies (< 1 Hz) and within the narrow bands of specific rhythms (and particularly of the 8–13 Hz alpha rhythm) that appear and disappear somewhat randomly in time. Signals collected on two or more electrodes exhibit changing levels of correlation, due either to physical proximity (that is, sharing of immediate influences from the cortical surface) or to actual coordination between different cortical sites, thus reflecting shared neuron activity within the brain itself. Spectral content and correlation have been related to various emotional and behavioral states.

Imbedded in this sustained "spontaneous" or "ongoing" electrical activity, short, distinctive (0.5–2 sec) waveforms can be found that are evoked, for instance, when a brief sensory message (stimulus) such as a brief illumination of the visual field or a tap on the forearm is received by the subject. These "evoked responses" are small (a few microvolts) and somewhat buried in the ongoing activity. The characteristics of the stimulus determine the evoked potential waveform together with the stimulus "environment," such as the level of attention of the subject, the "expectation set," and the meaning of the stimulus in the context of the experiment.

Can these observable electrical brain signals be put to work as carriers of information in man-computer communication or for the purpose of controlling such external apparatus as prosthetic devices or spaceships? Even on the sole basis of the present states of the art of computer science and neurophysiology, one may suggest that such a feat is potentially around the corner.

The Brain Computer Interface project, described later in this chapter, was meant to be a first attempt to evaluate the feasibility and practicality of utilizing the brain

1973



2020

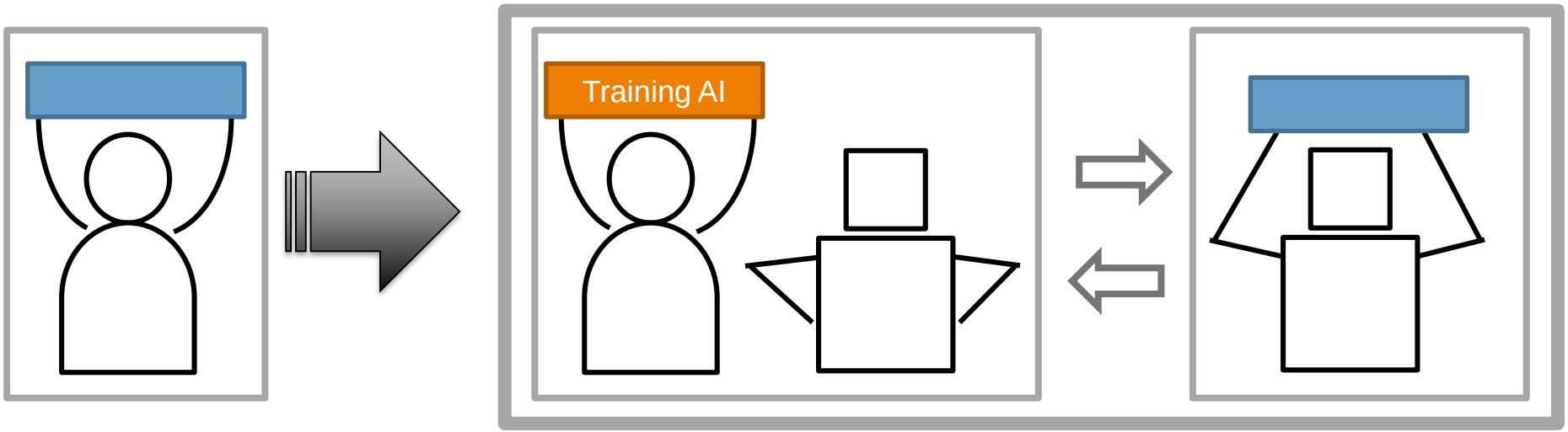
2045?

MEANINGFUL HUMAN CONTROL FOR HUMAN ENHANCER

Black mirror s3e5: Men against fire

Requires humans to have appropriate trust in their “sixth sense”

ROLE 4: AI AS A WORK CREATOR

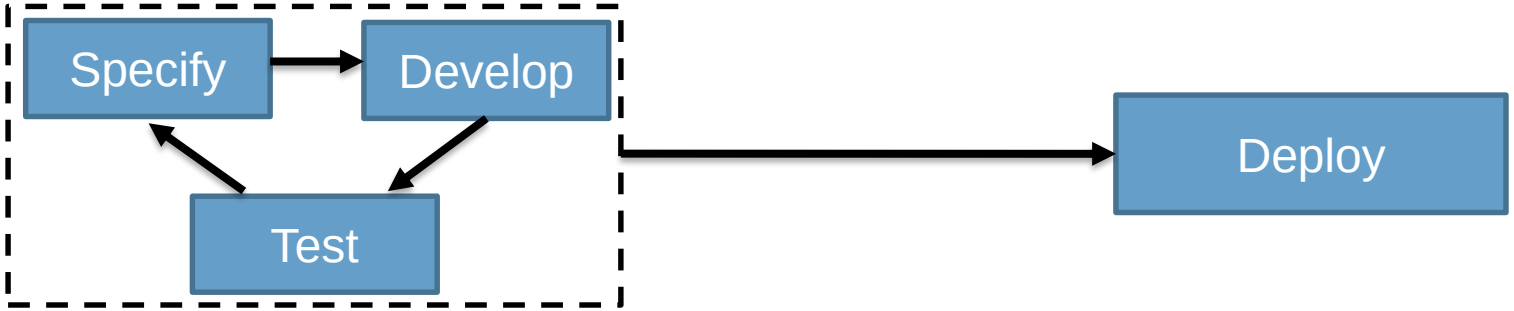


TOWARDS CONTINUOUS DEVELOPMENT

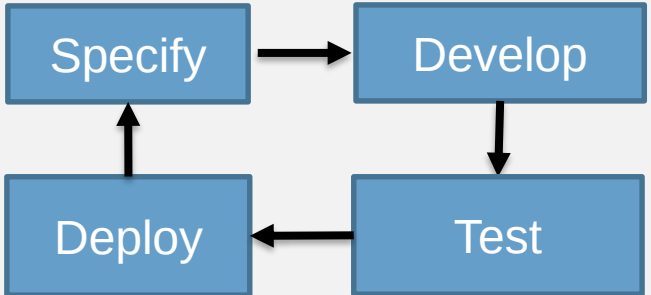
TNO innovation for life

Waterfall

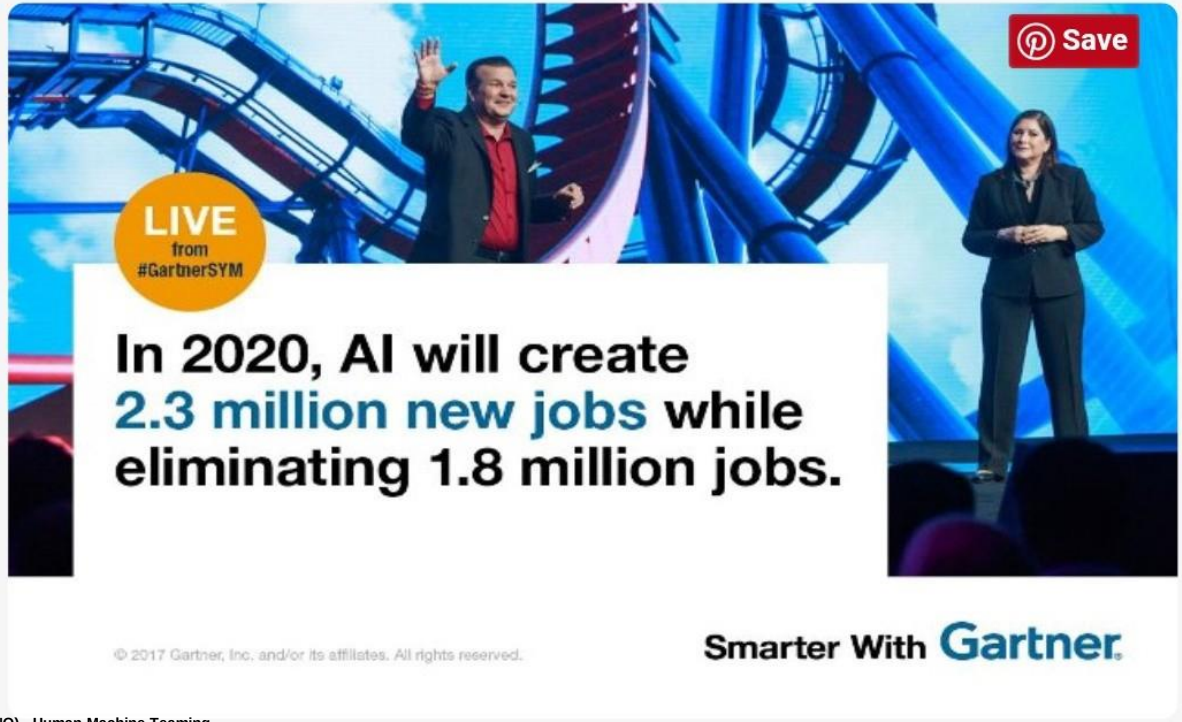
Agile



AI Systems Engineering



CURRENTLY


Save
LIVE
from
#GartnerSYM
In 2020, AI will create
2.3 million new jobs while
eliminating 1.8 million jobs.
© 2017 Gartner, Inc. and/or its affiliates. All rights reserved.
Smarter With Gartner

MHC BY AI ENGINEER

- MHC is continuously exercised by
  - Training Machine Learning components
  - Simulation (red teaming)
  - Control engineering
  - …
- Value-sensitive design

(a)

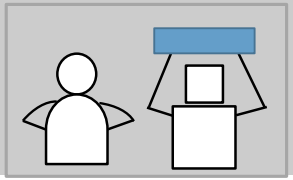
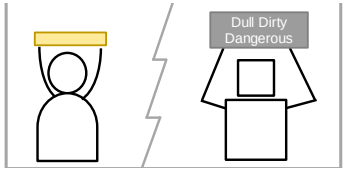
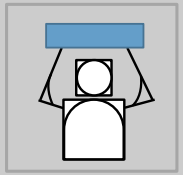
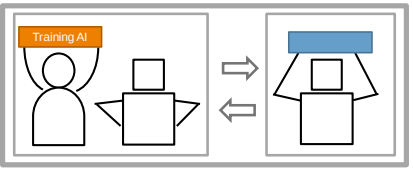
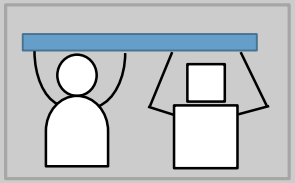
(b)

(c)

(d)

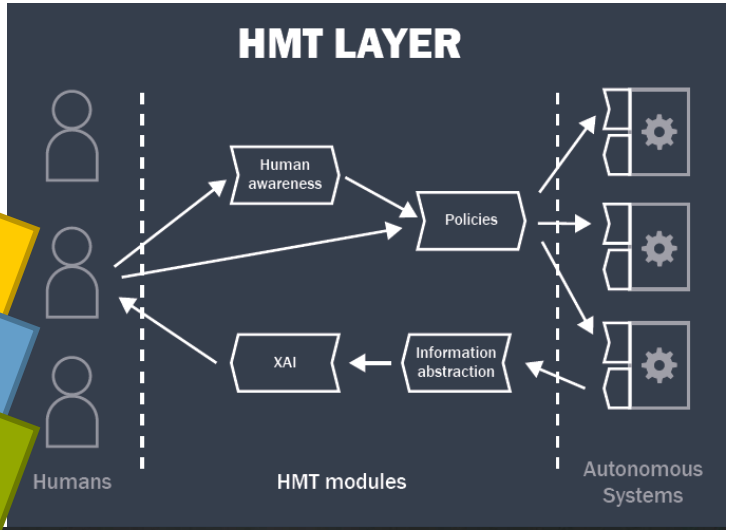
(e)

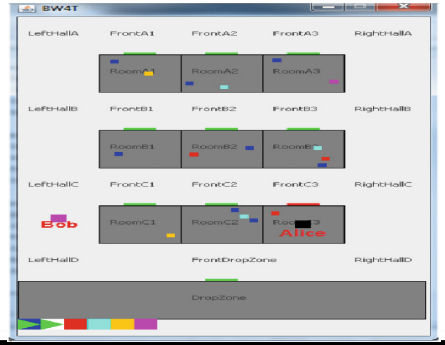
(f)

Pattern	Role of AI	Task properties	AI consequences	Human consequences
	Substitution	Ethically non-controversial	AGI or higher	Becomes irrelevant
	Obedient servant	Predictable	Must be robust	Deskilling
	Human Enhancer	Complex	Integrable with biology	Becomes upgraded
	Job creator	Complex	Must be continuously updateable	Must understand AI technology
	Teampartner	Complex	Must have teaming skills, s.a. XAI, TOM.	Must become team partner

HMT AT TNO

TNO innovation for life

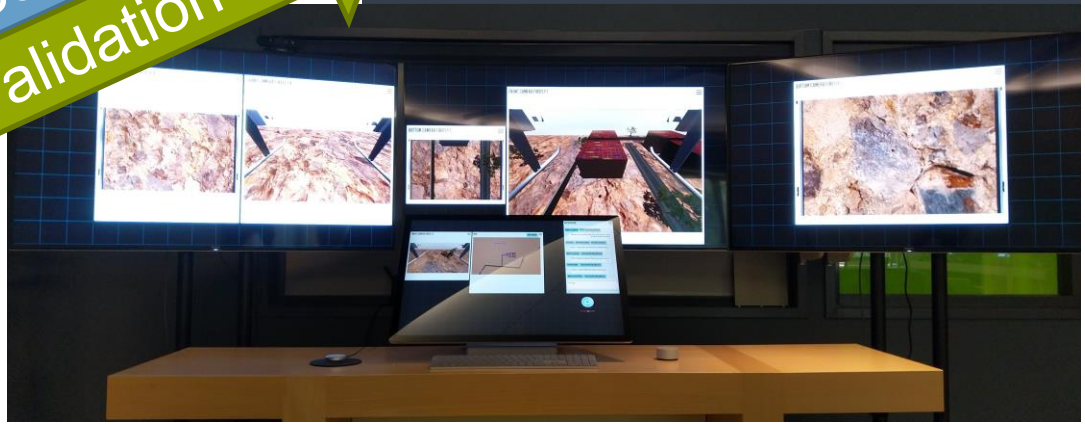




HMT specification

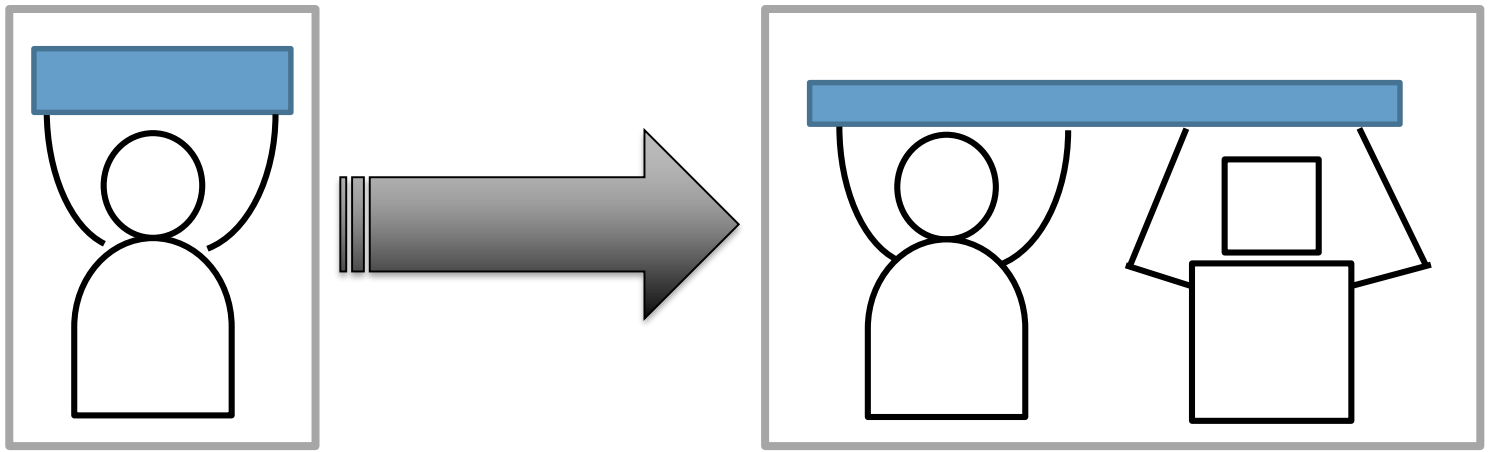
HMT technology

HMT validation

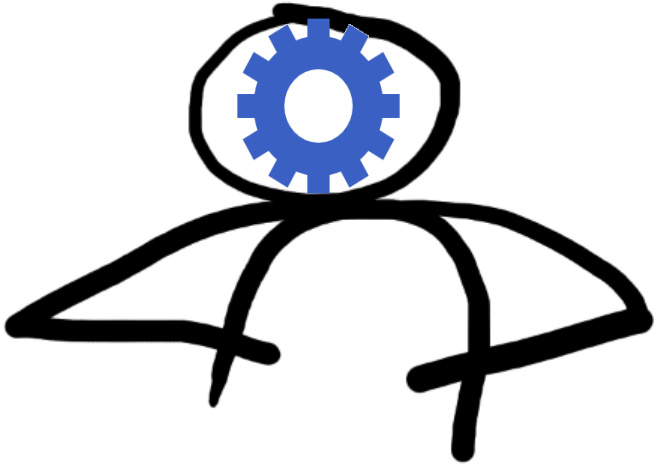


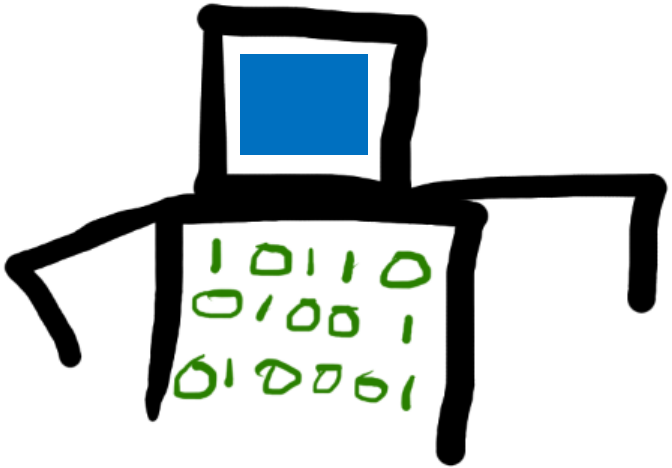
XAI

AI AS A TEAMPARTNER

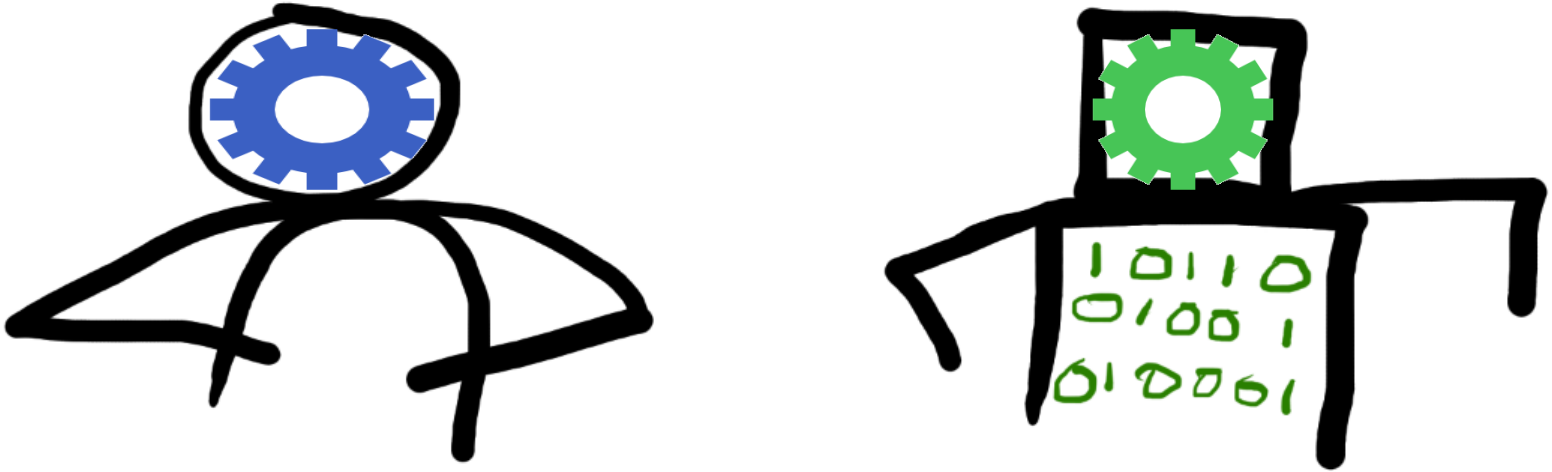


COMPUTER AS A TOOL

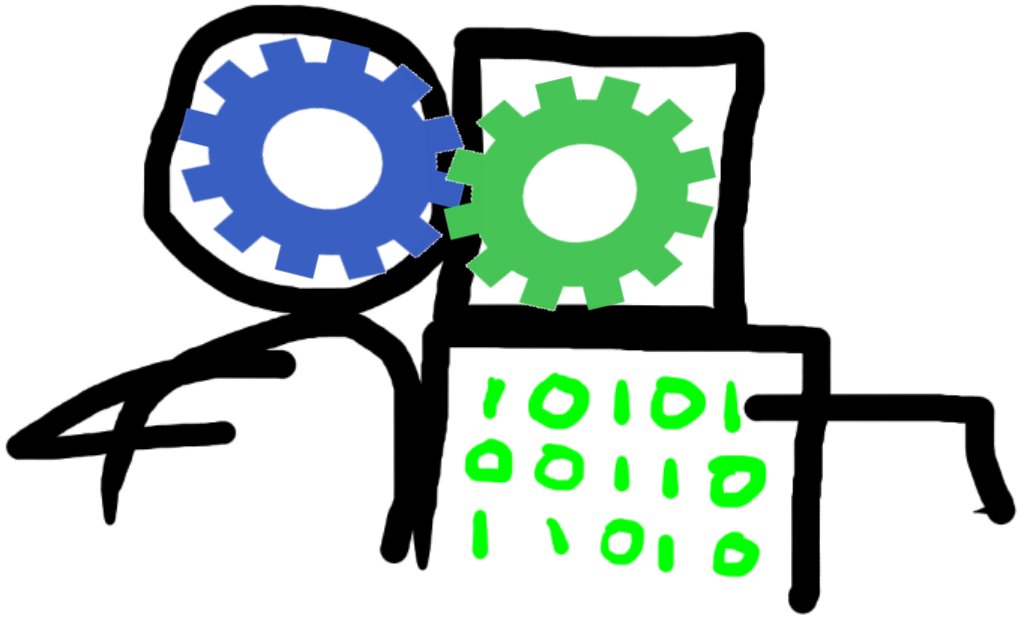




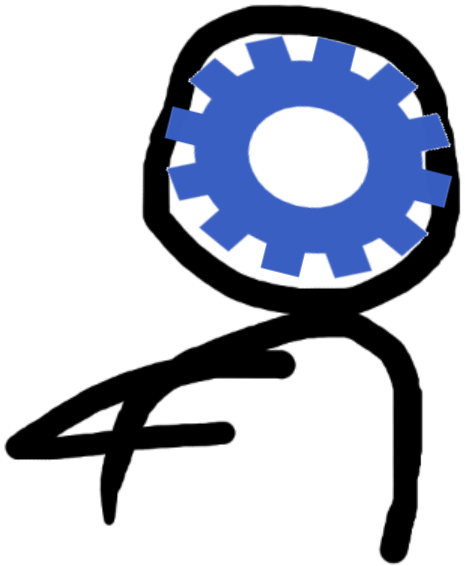
COMPUTER AS AN ISOLATED AGENT

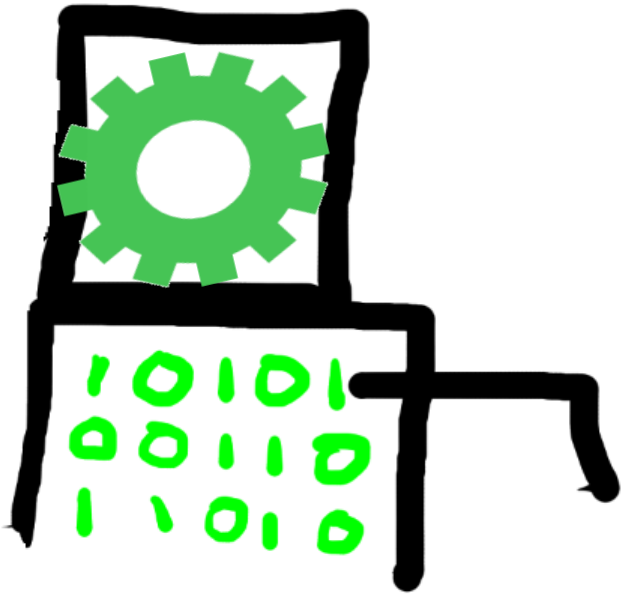


COMPUTER AS A TEAMMATE

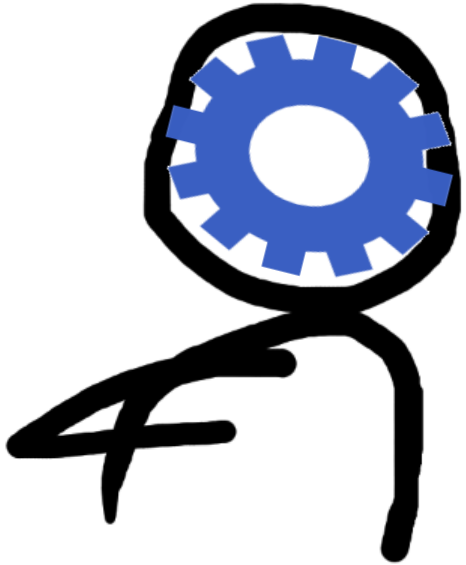


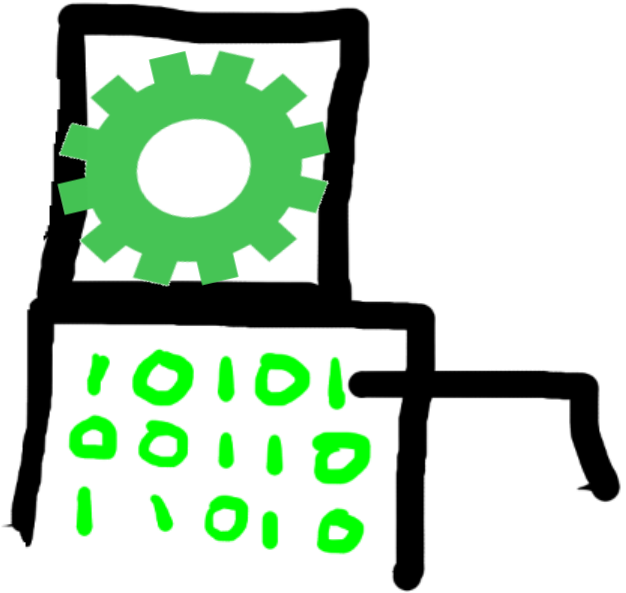
TEAMING HAS A TEMPORAL DIMENSION



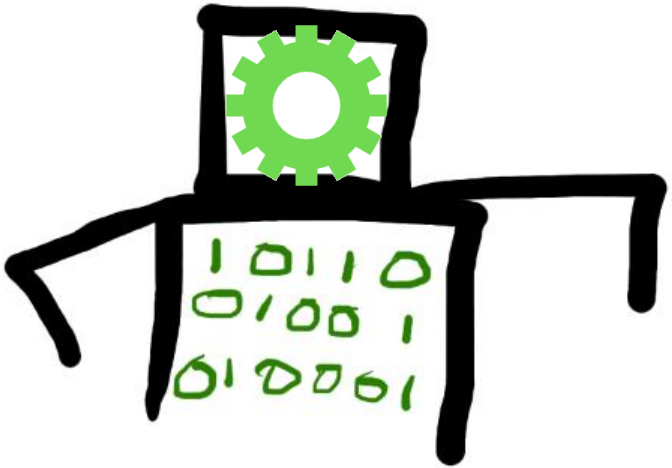


MACHINE INITIATIVE



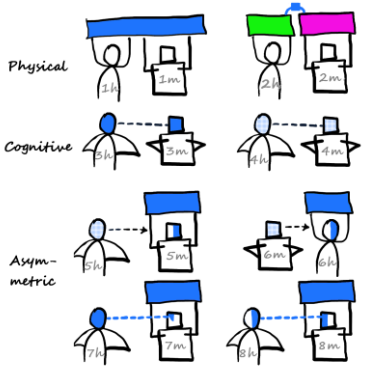


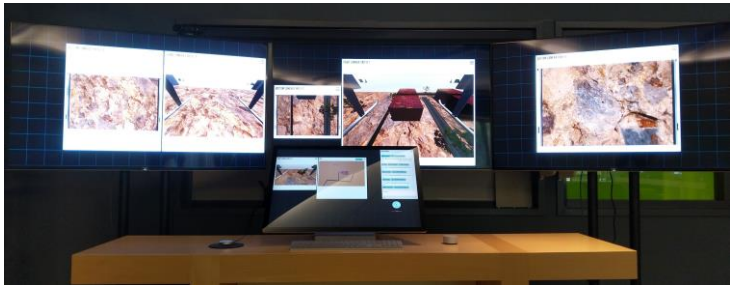
RUNAWAY AI

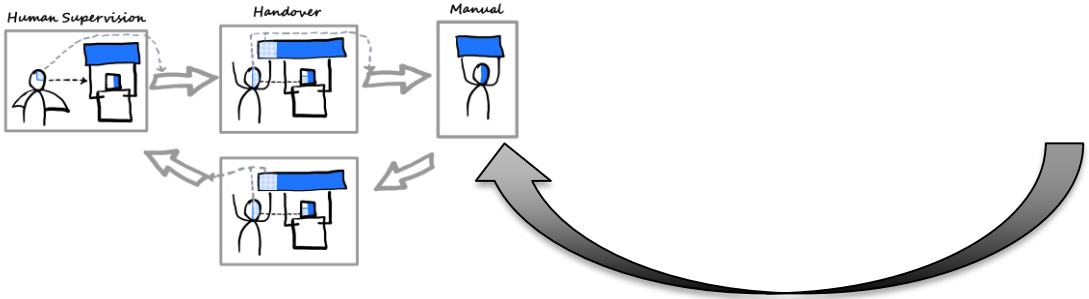


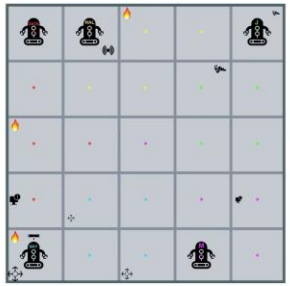
Specify

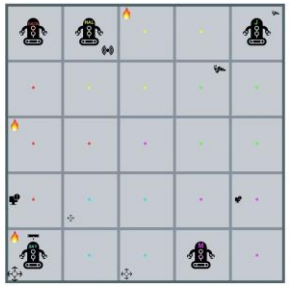
Test





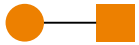




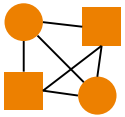


SCALES OF HUMAN MACHINE TEAMS

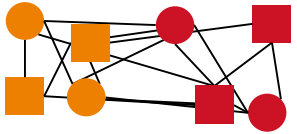
Machine in isolation



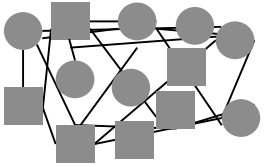
Human Machine Dyad



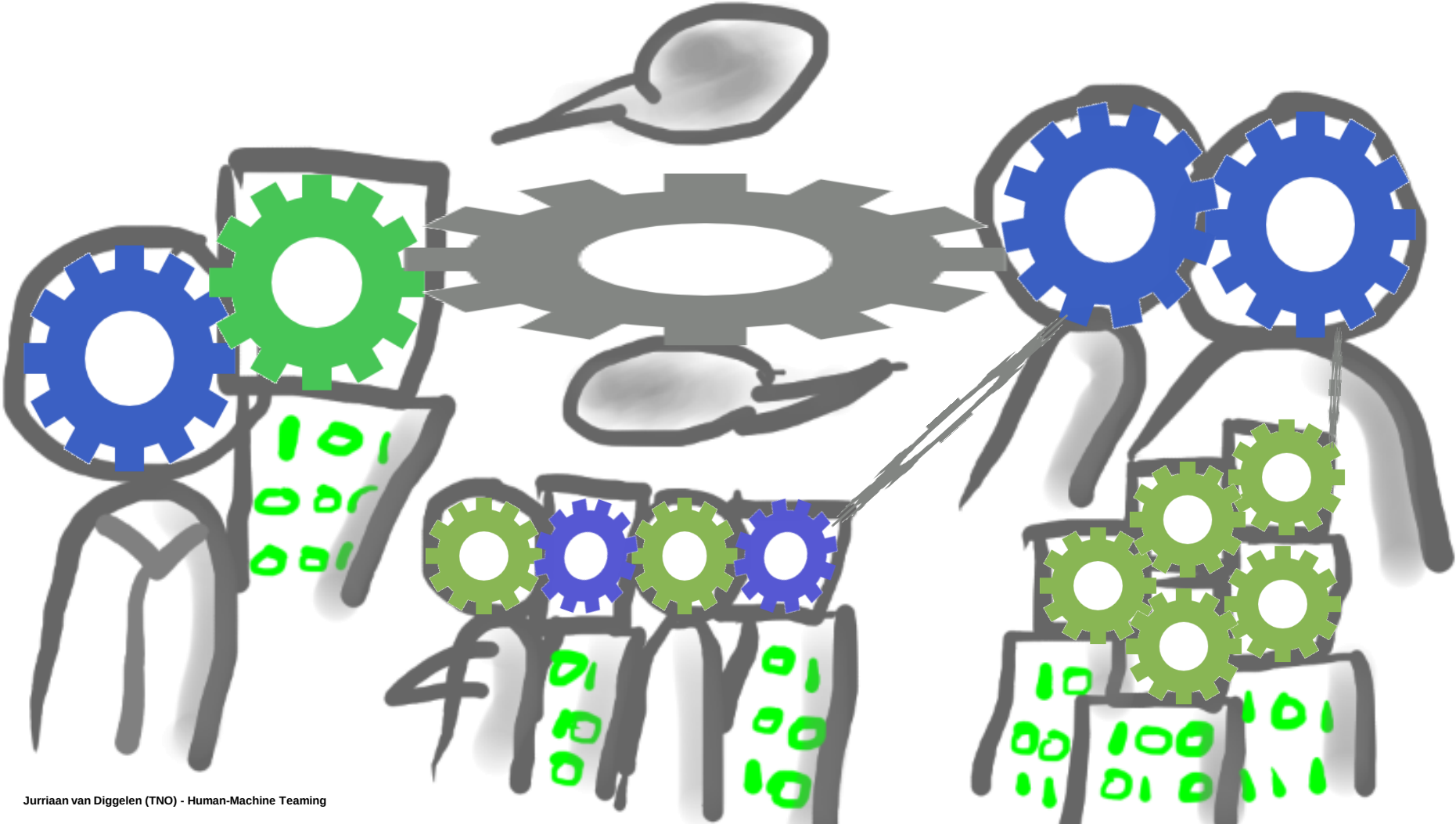
Human Machine Team



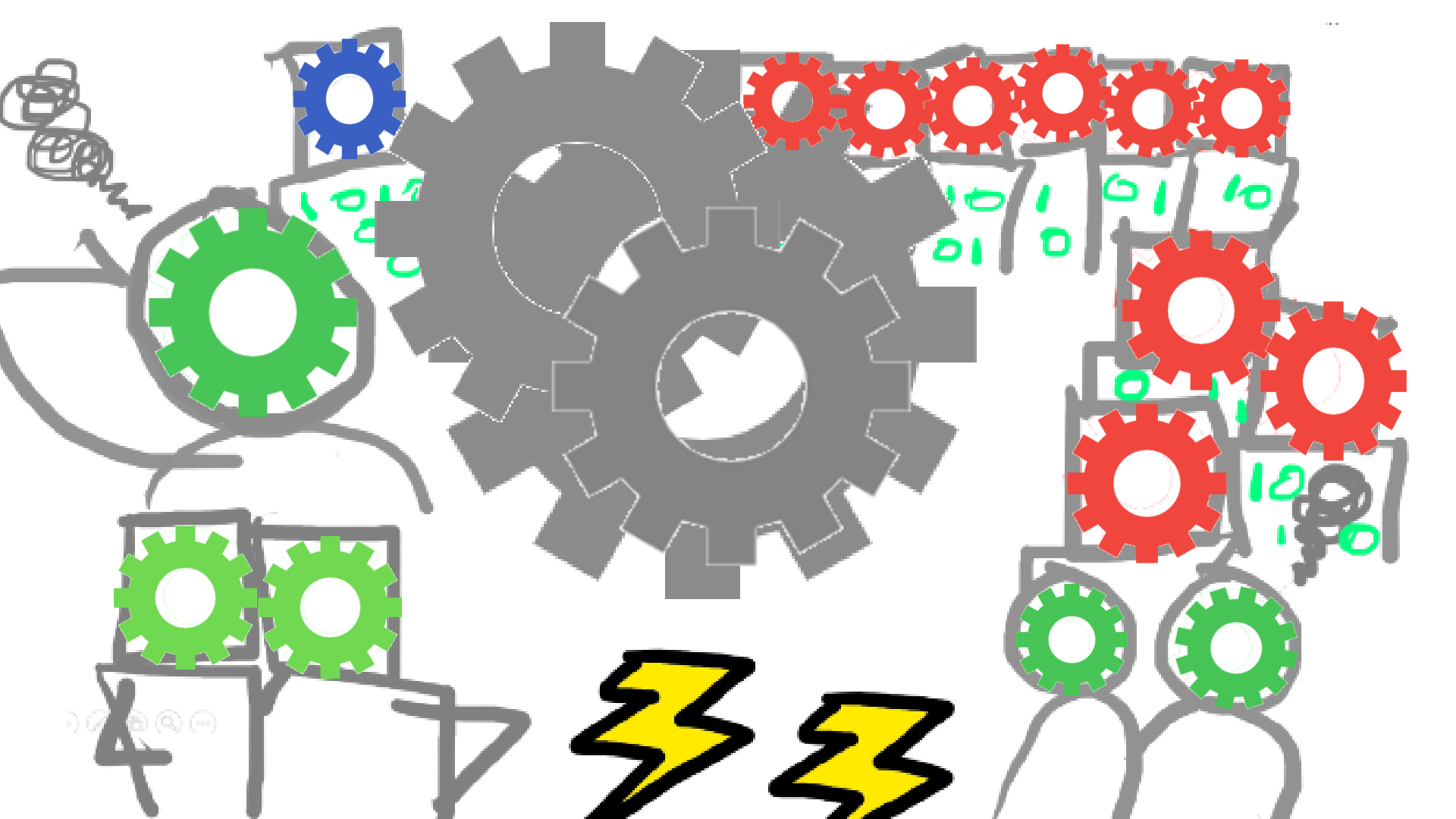
Hybrid Multi-Team System

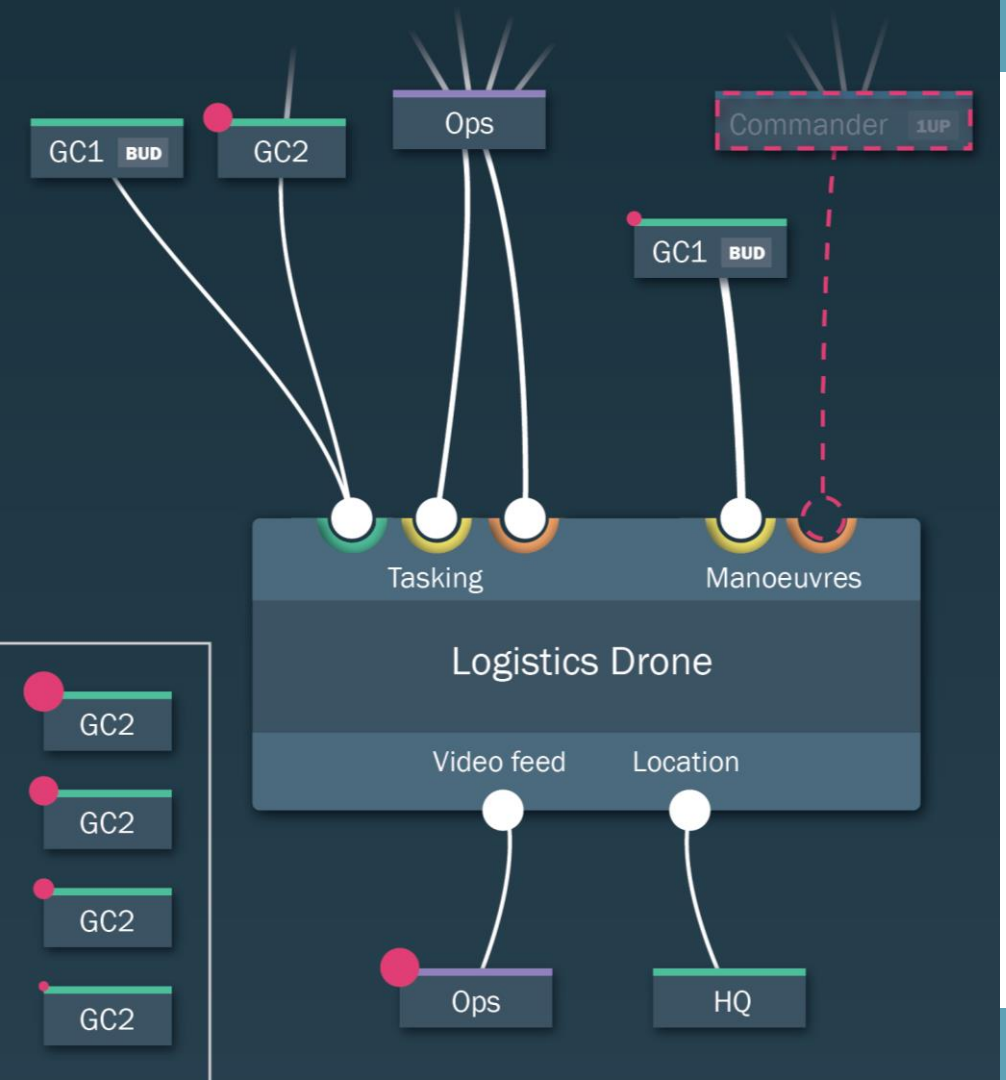


Hybrid Collective



Jurriaan van Diggelen (TNO) - Human-Machine Teaming





PROTO DESIGN PATTERN

GLANCEABLE VIEW

DESIGN PROBLEM

The Interdependency Inspector is not meant to be monitored continuously. In fact, the user will often only the Interdependency Inspector when alerted, when something has gone wrong or when something unexpected has happened. Therefore, the Inspector needs to be clear about what has recently changed in the overview and make important aspects salient to the user.

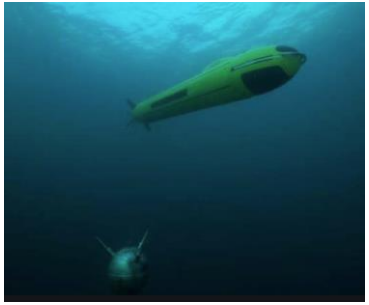
DESIGN SOLUTION

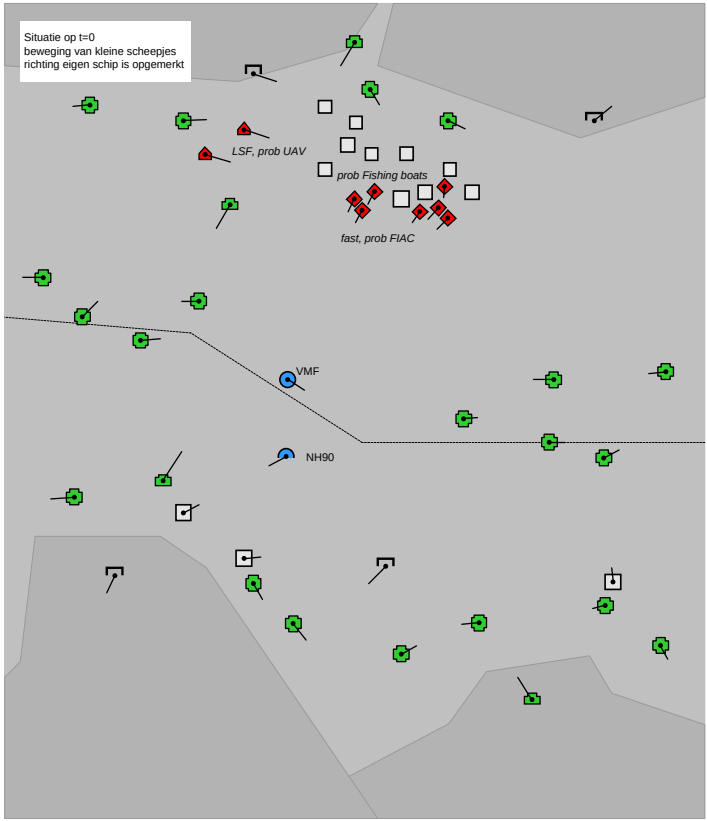
Salient circles are used to mark the changes in interdependency relations. Each change is marked with a circle, which decreases in size as time passes. Thus, the size of the circle serves as an indication of the recency of the change.

Missing relations that were previously connected, are marked with the same color.

APPLICATIONS OF MMT FOR NAVY

Jurriaan van Diggelen (TNO) - Human-Machine Teaming



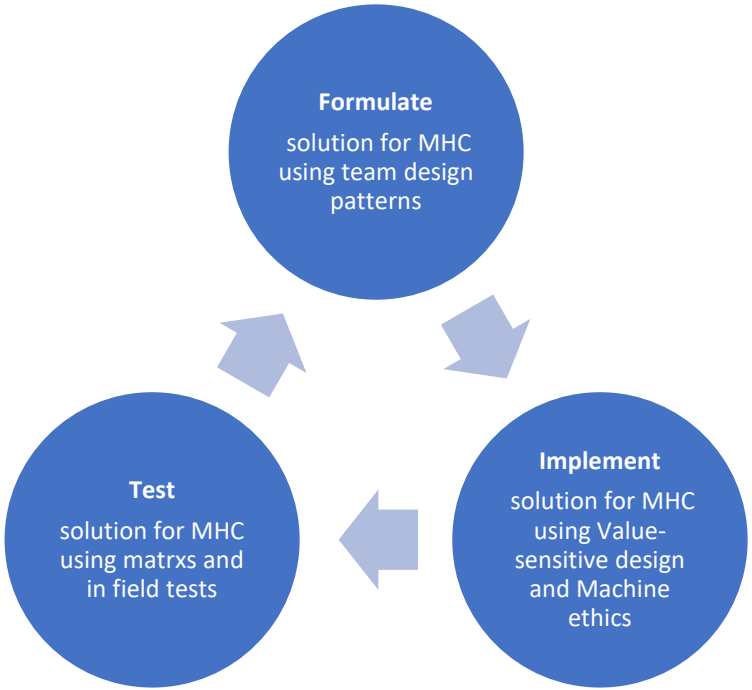


FIELDTESTS

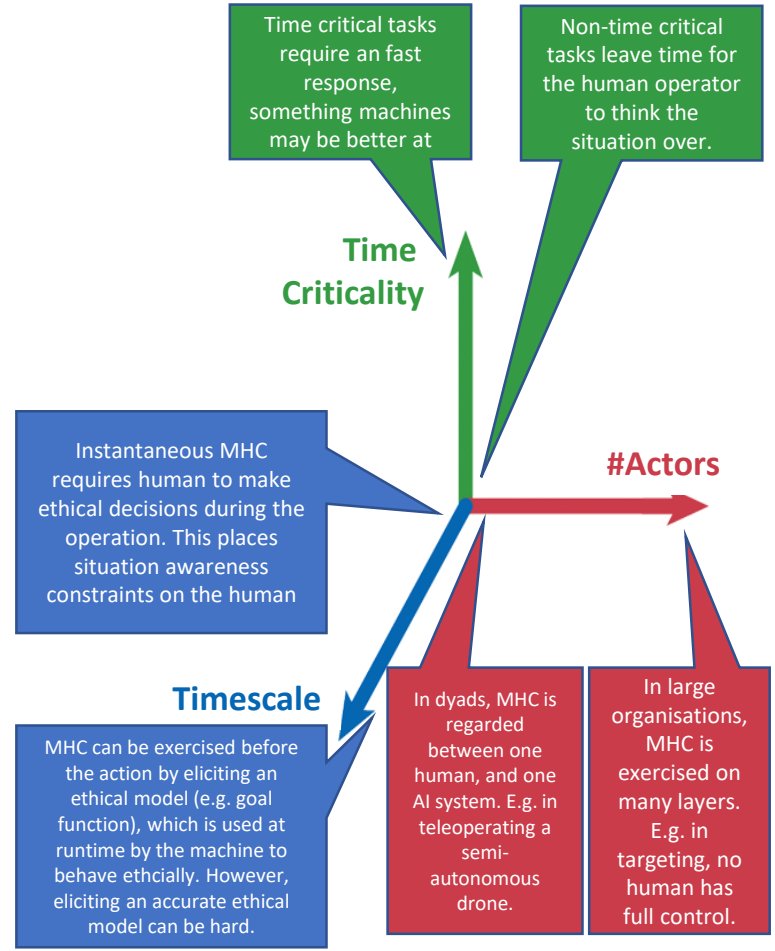
Goal of the project

To provide concrete, validated directives to Dutch defense for developing Meaningful Human Control over AI-based and autonomous systems in human-agent teams

Meaningful Human Control in Human Agent Teams



Dimensions of Meaningful Human Control



"The winner of the robotics revolution will not be who develops this technology first or even who has the best technology, but who figures out how to best use it."

Paul Scharre,
Robotics on the Battlefield Part 1: Range, Persistence and Daring

ARTIFICIAL INTELLIGENCE
NATURAL INTELLIGENCE
QUESTIONS?
SUMANTA BARUAH