UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT
DROIT D'AUTEUR

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document. Toute contrefaçon, plagiat ou reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10

UNIVERSITÉ DE LORRAINE
34 COURS LÉOPOLD — CS 25233
54052 NANCY CEDEX
TÉL : 03 54 50 54 00
CONTACT@UNIV-LORRAINE.FR
WWW.UNIV-LORRAINE.FR

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Fondamentaux du Jumeau Numérique Humain pour la Collaboration Humain-CPS

Human Digital Twin Fundamentals for Human-CPS Collaboration

THÈSE

présentée et soutenue publiquement le 9 juillet 2026

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention Automatique, Traitement du signal et des images, Génie informatique)

par

Ben Gaffinet

Composition du jury

Présidente :	Hind Bril El Haouzi	Professeure, Université de Lorraine
Rapporteurs :	Damien Trentesaux	Professeur, Université Polytechnique Hauts-de-France
	Caterina Rizzi	Professeure, University of Bergamo
Examinateurs :	Kurosh Madani	Professeur, Université Paris-Est Creteil
	Esma Yahia	MdC, HDR, Arts et Métiers Sciences et Technologies
	Christoph Stahl	Senior R&T Associate, Luxembourg Institute of Science and Technology
Directeurs de Thèse :	Hervé Panetto	Professeur, Université de Lorraine
	Yannick Naudet	Lead Scientist, ADR, Luxembourg Institute of Science and Technology

Université de Lorraine, CRAN, CNRS — UMR 7039

Mis en page avec la classe thesul.

Acknowledgments

In a different life, not that long ago, private sector work drove me from demo to demo, things had to be up and running fast. Along the way, I always regretted not having the time to dig deep into open scientific questions that I stumbled on periodically. That formed the core motivation to pursue a PhD; for once dig deep and go beyond the currently existing knowledge, leaving a tiny mark on the vast tapestry of human knowledge. Now, three years later, the result is here, my small contribution is made. But, the journey was a thousand times more valuable. Somehow, I held the expectation that this would be like any other job in Research and Development, completely overlooking the *student* part in PhD student. Yes the result matters, but the real challenge was learning the profession of research. And as with any profession, learning requires honest effort and great mentorship. The road leading to this manuscript was definitely *not* a lonely one. It is the result of a three-year long collaboration, where supervisors, colleagues, and fellow student shaped my thinking through rigorous discussions and casual coffee breaks. But, also the challenge of talking about work with friends and family left their mark. Nothing is more humbling than the inability to explain one's PhD topic their mother. I feel grateful for the support I had in my private life and at work throughout the years.

I was blessed to have two supportive and highly available supervisors on my team for this journey. To Yannick Naudet, my PhD co-director; you have my deepest thanks for your endless rigour and unconditional availability to discuss and challenge any idea, experiment or paper draft I may have. Your commitment has been unwavering and influenced many of my decisions for the better. I am convinced the thesis would never have reached conceptual clarity without your continuous challenge. To Hervé Panetto, my thesis director; I am grateful for your availability and guidance throughout this PhD. Despite the distance between Luxembourg and Nancy you always managed to help me navigate the challenges of both research and administration. It was a great pleasure to welcome you at our wedding in India, your presence was the best gift, and my family still asks when they will meet you again.

To Jana Al Haj Ali; As my fellow PhD colleague & friend, you have always been one step ahead and relentlessly supportive when I was facing the same challenges you just overcame. It benefited me greatly, helping me to keep calm during hard times. Being on the same project, working on similar problems with different perspectives provided fertile ground for all our discussions and exchanges over the years. With both our PhDs coming to an end I hope that our paths cross again and a new era of discussion, collaborating and co-authoring can start.

I thank my *Thesis Jury*; Hind Bril El Haouzi, Damien Trentesaux, Caterina Rizzi, Kurosh Madani, Esma Yahia, and Christoph Stahl. Your diverse perspectives challenged the content of this thesis while respecting its scope and limitations. I enjoyed every moment of our discussion and thank you for the time you took to engage with my work. A special thanks goes to my *Comité de Suivi de Thèse*, Hind Bril El Haouzi and Christoph Stahl; I thank you for our yearly meetings where your external point of view. It opened new perspectives and reflections that clearly enriched this thesis.

A big thank you to the whole LIST team for providing a healthy and productive work environment. You have been dependable in all aspects, and I could always count on your support when needed. I also wish to thank to CRAN laboratory. My physical presence was rare, but anytime I had the chance to be in Nancy I felt welcome and as part of the team.

Last but certainly not least, to my friends & family; I thank you with all my heart for being in my life. During this PhD process you've been a much needed source of stability and normalcy. But foremost, you've shaped me into who I am today. To my wife Priyanka; you've been a lifeline for long stretches of this thesis, especially towards the end. It was a heavy burden. Without a doubt I can say that I would not have made it without you. I can never express how grateful I am to share my life with you, through all the good and the bad times alike.

Acknowledgments
This work has been partially supported by the ANR French National Research agency and the FNR Luxemburgish National Research funds project AI4C2PS (INTER/ANR/22/17164924/AI4C2PS), 2023-2025.

As iron sharpens iron,
so one person sharpens another.
Proverbs 27:17

Abstract

Industry 5.0 envisions a transformative shift from system-centric automation toward a human-centric symbiosis between workers and Cyber-Physical Systems. In this paradigm, effective collaboration requires machines to move beyond reacting to physical movements and develop an awareness of unobservable human cognitive states, such as intentions, knowledge levels, and mental fatigue. Human Digital Twins (HDTs) represent a key enabling technology for this vision, yet their development has been hampered by conceptual ambiguity and the inherent difficulty of modelling human cognition.

This thesis establishes the conceptual foundations for HDTs and investigates their potential to enable adaptive, human-aware collaboration within industrial contexts. The research follows a structured path across four primary contributions. First, it proposes a general domain-agnostic definition and classification system for HDTs, clarifying the unique one-to-one relationship between a human and their digital counterpart as well as the required level of data integration. Building on the essential elements of HDTs, we develop a meta-model anchored in General Systems Theory (GST), providing a unified framework that delineates the boundaries and interfaces between digital, physical, and cyber-physical entities. Moving towards an application case in Human-CPS Collaboration, the thesis proposes a decentralized DT-based architecture designed to achieve cognitive interoperability, which is the alignment of perceptions and sharing of intentions between collaborating agents. Lastly, the research formalises an emulation model for human cognition and experimentally evaluates an ACT-R based instantiation. Two experiments involving an assembly task demonstrate ACT-R's performance and reveals its limitations as emulation model. The first experiment successfully reproduces human-like learning and forgetting trends. The second experiment tests the effectiveness of real-time synchronisation to improve response-to-response accuracy. However, the findings also point to significant limitations caused by individual-differences including diverse mental strategies, and the challenge of inferring cognitive states from observable behaviour. This highlights the need for richer synchronisation signals and procedures in future high-fidelity HDTs.

Keywords: Human Digital Twin, Human-CPS Collaboration, Cognitive Emulation, Cognitive Architecture, Industry 5.0

Résumé

L'Industrie 5.0 envisage une transition majeure, passant d'une automatisation centrée sur les systèmes à une symbiose centrée sur l'humain entre les travailleurs et les systèmes cyber-physiques (CPS). Dans ce paradigme, une collaboration efficace exige que les machines aillent au-delà de la simple réaction aux mouvements physiques et développent une conscience des états cognitifs humains non observables, tels que les intentions, les niveaux de connaissance ou la fatigue mentale. Les jumeaux numériques humains (JNH) constituent une technologie clé pour cette vision, mais leur développement a été entravé par une ambiguïté conceptuelle et la difficulté inhérente à la modélisation de la cognition humaine complexe.

Cette thèse pose les fondements conceptuels des JNH et explore leur potentiel pour permettre une collaboration adaptative et sensible à l'humain dans des contextes industriels. La

recherche suit un parcours structuré articulé autour de quatre contributions principales. Elle propose une définition générale indépendante du domaine et un système de classification pour les JNH, clarifiant les exigences en matière d'intégration automatisée des données et la relation unique de type « un-à-un » entre un humain et son homologue numérique. En nous appuyant sur les éléments essentiels des HDT, nous développons un méta-modèle systémique ancré dans la théorie générale des systèmes, offrant un cadre unifié qui définit les limites et les interfaces entre les entités numériques, physiques et cyber-physiques. En mettant l'accent sur la collaboration entre l'humain et les (CPS), cette thèse propose une architecture décentralisée basée sur les jumeaux numériques (JN), conçue pour atteindre l'interopérabilité cognitive, c'est-à-dire l'alignement des perceptions et des intentions partagées entre les agents collaborant. Enfin, cette recherche formalise et évalue expérimentalement un modèle d'émulation de la cognition humaine utilisant l'architecture cognitive ACT-R. Deux expériences portant sur une tâche d'assemblage industriel démontrent la capacité d'ACT-R à reproduire les tendances d'apprentissage et confirment l'efficacité de la synchronisation en temps réel pour améliorer la précision de réponse à réponse. Cependant, les résultats identifient également des limites importantes causées par les différences individuelles, notamment la diversité des stratégies mentales, ainsi que la difficulté à déduire les états cognitifs à partir du comportement observable. Cela souligne la nécessité de disposer de signaux de synchronisation plus riches dans les futurs HDT haute fidélité.

Mots-clés: jumeau numérique humain, collaboration humain-cps, émulation cognitive, architecture cognitive, industrie 5.0

List of Publications

Journal Publications:

- Al Haj Ali, J., **Gaffinet, B**., Lezoche, M., Panetto, H., Naudet, Y. (2025). [Al Haj Ali et al., 2025a] Enabling human-CPS cognitive interoperability: Cognitive architectures as technologies for human-like cognitive digital twins, Journal of Industrial Information Integration, Volume 48, 2025, 100969, ISSN 2452-414X, doi:10.1016/j.jii.2025.100969.

- **Gaffinet, B**., Al Haj Ali, J., Panetto, H., Naudet, Y. (2025). [Gaffinet et al., 2025a] Human Digital Twins: A systematic literature review and concept disambiguation for industry 5.0, Computers in Industry, Volume 166, 2025, 104230, ISSN 0166-3615, `https://doi.org/10.1016/j.compind.2024.104230`.

- Al Haj Ali, J., **Gaffinet, B**., Panetto, H., Naudet, Y. (2024). [Al Haj Ali et al., 2024a] Cognitive systems and interoperability in the enterprise: A systematic literature review, Annual Reviews in Control, Volume 57, 2024, 100954, ISSN 1367-5788, `https://doi.org/10.1016/j.arcontrol.2024.100954`.

International Conferences:

- **Gaffinet, B.**, Al Haj Ali, J., Panetto, H., Naudet, Y. (2025). [Gaffinet et al., 2025b], Towards Cognitive Interoperability with Cognitive Human Digital Twins. 19th IFAC/IFIP International Workshop on Enterprise Integration, Interoperability and Networking, EI2N 2025, Oct 2025, Marbella, Spain. pp.332-349, https://doi.org/10.1007/978-3-032-15579-5_20.

- Al Haj Ali, J., **Gaffinet, B**., Panetto, H., Naudet, Y. (2025). [Al Haj Ali et al., 2025b], From Semantic Interoperability to Cognitive Interoperability: Enabling Human-CPS Collaboration in Industry 5.0. 6th IFAC/INSTICC International Conference on Innovative Intelligent Industrial Production and Logistics, IN4PL 2025, Oct 2025, Marbella, Spain. pp.103-122, https://doi.org/10.1007/978-3-032-15576-4_7.

- Naudet. Y., Zehnder, E., **Gaffinet, B**., Al Haj Ali, J., Panetto, H., (2025). [Naudet et al., 2025], A Human Ontology with Cognition and Ability Models for Human and Cognitive Digital Twins. 6th IFAC/INSTICC International Conference on Innovative Intelligent Industrial Production and Logistics, IN4PL 2025, Oct 2025, Marbella, Spain. pp.52-71, https://doi.org/10.1007/978-3-032-15576-4_4.

- Anastasiou, D., **Gaffinet, B.**, Naudet, Y. (2026). [Anastasiou et al., 2025] A Cognitive Social Robot in Manufacturing. In: Staffa, M., *et al.* Social Robotics + AI. ICSR+AI 2025. Lecture Notes in Computer Science(), vol 16131. Springer, Singapore. https://doi.org/10.1007/978-981-95-2379-5_44 .

- **Gaffinet, B.**, Naudet, Y., Panetto, H. (2025). [Gaffinet et al., 2025c] A Systemic Human Digital Twin Model for Human-Centric Systems. In: *IFAC-PapersOnLine*, *59*(10), 416-421, ISSN 2405-8963, https://doi.org/10.1016/j.ifacol.2025.09.072

- Naudet Y., Al Haj Ali, J., **Gaffinet, B**., Panetto, H. (2024). [Naudet et al., 2024], Cognition in digital twins for cyber-physical systems and humans: Where and why?. In: Dassisti, M., Madani, K., Panetto, H. (eds) Innovative Intelligent Industrial Production and Logistics. IN4PL 2024. Communications in Computer and Information Science, vol 2373. Springer, Cham. https://doi.org/10.1007/978-3-031-80775-6_27.

- Al Haj Ali, J., Lezoche M., Panetto, H., Naudet, Y., **Gaffinet, B.** (2024). [Al Haj Ali et al., 2024b] Cognitive architecture for cognitive cyber-physical systems, IFAC-Papers OnLine, 58(19), 1180–1185, ISSN 2405-8963, doi:10.1016/j.ifacol.2024.09.099.

- **Gaffinet, B**., Al Haj Ali, J., Panetto, H., Naudet, Y. (2023). [Gaffinet et al., 2023], Human-centric digital twins: Advancing safety and ergonomics in human-robot collaboration. In: Terzi, S., Madani, K., Gusikhin, O., Panetto, H. (eds) Innovative Intelligent Industrial Production and Logistics. IN4PL 2023. Communications in Computer and Information Science, vol 1886. Springer, Cham. https://doi.org/10.1007/978-3-031-49339-3_24.

- Bhattacharya, S., **Gaffinet, B.**, Naudet, Y., Panetto, H. (2026). [Bhattacharya et al., 2026] A Formal Framework for Evaluating Cognitive Emulation Models in Human Digital Twins. 23rd IFAC World Congress. Busan, South Korea.

- Kottagaha, K., Bokhorst, J.A.C., **Gaffinet, B.**, Emmanouilidis, C. (2026). An agent model abstraction for human-AI teaming cognitive coupling. 23rd IFAC World Congress. Busan, South Korea.

Contents

List of Figures **xiii**

List of Tables **xvii**

Acronyms **xix**

Résumé en Français (French Summary)

Introduction

1 Context and Motivation 1
1.1 Project context: AI4C2PS 3
2 Research Questions and Contributions 3
3 Structure 4

Chapter 1
State of the Art

1.1 Introduction 7
1.2 Prior Reviews 8
1.3 Descriptive Analysis 11
1.3.1 Timeline & Application Areas 11
1.3.2 Geographic Distribution 12
1.3.3 Concept of HDT 14
1.4 Challenges and Research Directions 18
1.4.1 Communication and Connectivity 18
1.4.2 Sensing Human Data 19
1.4.3 Modelling 19
1.4.4 Societal and Ethical Challenges 21
1.5 Discussion 22
1.6 Contribution Positioning 22

Chapter 2
Human Digital Twin Fundamentals

2.1 Introduction . . . 25
2.2 Classification & Definition . . . 25
2.2.1 HDT Specificities: Physical Object . . . 27
2.2.2 HDT Specificities: Data Flow to the Digital Object . . . 28
2.2.3 HDT Specificities: Digital Object . . . 28
2.2.4 HDT Specificities: Feedback to the Physical Object . . . 30
2.2.5 General Human Digital Twin Definition . . . 31
2.2.6 Special Case: Augmented Human Digital Twin . . . 31
2.2.7 Summary of Human Digital Twin Fundamentals . . . 33
2.3 Systemic Meta-Model . . . 34
2.3.1 Core Entities . . . 34
2.3.2 Systemic Grounding . . . 35
2.3.3 Human Digital Twin Meta-Model . . . 36
2.3.4 Human Meta-Model . . . 38
2.3.5 Sensors & HMI Devices . . . 39
2.3.6 Consolidated Systemic Meta-Model . . . 41
2.3.7 Human Digital Twin Functions . . . 41
2.4 Conclusions . . . 43

Chapter 3
Human Digital Twins for Human-CPS Collaboration

3.1 Introduction . . . 45
3.2 Challenges in Human-Robot Collaboration . . . 46
3.2.1 General HRC Challenges . . . 47
3.2.2 Human Digital Twins in HRC . . . 48
3.2.3 Cognitive Interoperability . . . 50
3.3 Digital Twin Based Architecture . . . 52
3.3.1 Entities and Data Flows . . . 53
3.3.2 Use Cases . . . 57
3.3.3 Challenges . . . 61
3.4 Emulation Model for Human Cognition . . . 62
3.4.1 Emulation Model Formalisation . . . 62
3.4.2 Cognitive Modelling . . . 65
3.4.3 Cognitive Architectures . . . 67
3.5 Conclusions . . . 71

Chapter 4
Experiments

4.1 Introduction . . . 73
4.1.1 Structure . . . 75
4.2 ACT-R Theoretical Background . . . 75
4.2.1 ACT-R General Structure . . . 75
4.2.2 Declarative Memory . . . 78
4.2.3 Perception and Motor Modules . . . 81
4.2.4 Personalisation . . . 82
4.3 Simulation and Emulation . . . 84
4.3.1 Simulation with ACT-R . . . 85
4.3.2 Emulation with ACT-R . . . 85
4.4 Experiment 1 - Assembly Task Simulation . . . 87
4.4.1 Task Description . . . 87
4.4.2 Model Description . . . 88
4.4.3 Simulation Results . . . 91
4.4.4 Discussion and Conclusions . . . 93
4.5 Experiment 2 - Retrieval Behaviour Emulation . . . 95
4.5.1 Calibration Phase . . . 96
4.5.2 Exploration Phase . . . 103
4.6 Conclusions . . . 116

Chapter 5
Conclusions & Research Directions

5.1 Thesis Summary . . . 117
5.2 Perspective and Research Directions . . . 118

Appendix A: Systematic Literature Review Categorisation Table

Bibliography

List of Figures

1 Structure of the thesis and contributions for each chapter. 4

1.1 Timeline of Human Digital Twins (HDT) and various conceptual perspectives. . . 11
1.2 Number of publications by year and by application area. In 2023 papers were screened until the 15th of June. 12
1.3 (a) Total number of publications by application area. (b) Total number of papers categorised as "Other" and number of unique categories that have less than 4 papers. 13
1.4 Number of publications by year and continent. 13
1.5 Number of publications by country and application area. 14
1.6 **(a)** Digital Model, **(b)** Digital Shadow and **(c)** Digital Twin as introduced in Kritzinger et al. [2018]. The categories are distinguished by the level of data integration. 15
1.7 Level of data integration in reviewed paper, categorised by application domain and whether the study is conceptual in nature or a practical case-study. 17
1.8 Count of publications categorised by the nature of the twinned entity. 18

2.1 Vocabulary for Digital Twin and Human Digital Twin. 26
2.2 Overview of entities and data flows for ***(i)*** Human Digital Model; ***(ii)*** Human Digital Shadow; ***(iii)*** Human Digital Twin and; ***(iv)*** traditional Digital Twin. . . 27
2.3 Data flows of an augmented Human Digital Twin (aHDT) where a human and a strongly coupled technical system are twinned to a single DT. 32
2.4 Essentials of Human Digital Twins and categorisation framework based on the twinned entity and level of data integration 34
2.5 The *Human Digital Twin - Human Individual* pair, and the essential entities to enable emulation of the human by the HDT and interaction between both. 35
2.6 The four essential elements for HDT development as different types of systems, and the main elements of a systemic meta-model. 36
2.7 Meta-Model for Human Digital Twin. 37
2.8 Human model: the Emulation Model of the Human Individual, as a composition of specialised models of the HDT's twinned entity. 38
2.9 Meta-Model of a *Human Individual* as a *Physical System*, including the main sub-systems and interaction with the *Environment*. 39
2.10 Consolidated meta-model for the *Human Digital Twin*, *Human Individual* including the information transfers by both *Sensors* and *HMI Devices* 40
2.11 Human-HDT architecture with its main functions. 42

3.1 Conceptual architecture for Human-Robot collaboration from the AI4C2PS project. 46

3.2 Simplified illustration of the entities in an HRC setup and their Digital Twins or Digital Shadows respectively. . . . 53
3.3 Classification of actions in multi-agent systems based on the formalisation presented in [Weyns and Holvoet, 2004] . . . 54
3.4 Digital Twin based architecture and data flows for HRC scenarios. . . . 55
3.5 Basic input-output relation of an emulation model for human cognition in relation to the physical environment. . . . 63
3.6 Relation between the environment, model and various processes. . . . 64
3.7 Common Model of Cognition from Laird et al. [2017]. . . . 68

4.1 Experiment overview and connection to HDT development. . . . 74
4.2 Structure of ACT-R and mapping to specific brain structures from Ritter et al. [2019] . . . 75
4.3 Coloured squares represent buffer states. Cognitive cycle of ACT-R in three steps **(1)** *matching* refers to reading the buffer states, **(2)** *selection* of production rule, **(3)** *execution* of production rule leading to a change in buffer state. Asynchronous execution of modules can lead to further concurrent changes in buffers. . . . 76
4.4 Figure showing the interactions between goal, procedural memory and declarative memory. Adapted from [Lovett et al., 1999]. . . . 78
4.5 Example of activation spreading. A memory chunk encodes the location of Lego bricks that are defined by their size and colour. In the example the goal is to find the correct location of a small red brick. . . . 80
4.6 Interaction between ACT-R and environment: (1) perception of stimuli, (2) internal processing and action selection, (3) execution of motor actions. . . . 85
4.7 Setup for the assembly task and the two product types to assemble. . . . 88
4.8 Overview of buffers and chunk types of the ACT-R model. Arrows indicate which chunktype is accepted by which buffer. . . . 90
4.9 Flowchart of the model behaviour. Yellow boxes represent physical actions that change the state of the environment. . . . 91
4.10 Assembly performances: (a) Time spent to finish each product. The colours indicate the product type (red for duck, blue for gorilla). (b) Number of drawers opened to find the required component for one assembly step. Each shaded area corresponds to one product. . . . 92
4.11 Overview of the experiment phases, tasks, ACT-R model use, and objectives. . . 96
4.12 Sequence in which digits are presented on screen during the digit working memory task in the calibration phase Figure adapted from Lovett et al. [1999]. . . . 98
4.13 Aggregate data of observed performances compared to the original trends in Lovett et al. [1999]. . . . 98
4.14 Major production rules for the calibration model and the . . . 99
4.15 Simulated success rates compared to the observed mean success rates. The base model produces a recall success rate that exceeds human performance, while the aggregate model approaches the observed behaviour. . . . 101
4.16 Drawer setup, drawer contents and structure of the exploration experiment. . . . 104
4.17 Model functioning overview. ACT-R is managing the declarative memory and retrievals. A customized Environment module is handling stimuli and executing actions. . . . 106
4.18 (a) Observed success rates for each round of questions. (b) Reported score on the single ease question (SEQ) . . . 109

4.19 Simulated success rates of the base model (green), aggregate model (blue), in comparison with the observed mean success rate (dashed red). 110
4.20 Relation of W parameters and: (a) change in F1 score between personalised and regular model, (b) observed exploration success rates. 111
4.21 (a) F1 score for aggregate and synchronised model compared to a random predictor model. (b) Recall success rate for synchronised and aggregate model compared to observations. 112

List of Tables

1.1 Classification of prior reviews and positioning of our own review. 9
1.2 Number of publications grouped by level of integration and whether they provide a definition before using the HDT concept. 16

3.1 Use Case 1: Potential benefits of the DT-based collaborative assembly setup compared to the non-DT based alternative. 59
3.2 Use Case 2: Potential advantages of an HDT-based guidance system during an assembly task compared to a non-HDT based alternative. 60
3.3 Comparison of shortlisted cognitive architectures as they fulfil emulation requirements. (R1: biological plausibility, R2: environment interaction, R3: interpretable, R4: library for integration) 71

4.1 ACT-R declarative module tuneable sub-symbolic parameters 81
4.2 ACT-R Vision module parameters 82
4.3 Changes in environment module and ACT-R model when an action is resolved. . 89
4.4 Success rate of first retrieval attempt for each product assembly. 93
4.5 Parameters for the calibration module. 100
4.6 Personalised parameters and model performances. 102
4.7 Metrics used for exploration phase. 106
4.8 Reinforcements of chunks during *recall location* instructions depending on retrieval outcome. 107
4.9 Parameters for the exploration model. See Table 4.1 in section 4.2.2 for the full description of the parameters. 108
4.10 Impact of personalisation on performance metrics. 110
4.11 Impact of synchronisation on performance metrics 111
4.12 Main observations from post-task interview with participants. 113

Acronyms

ACT-R: Adaptive Control of Thought—Rational

aHDT: augmented Human Digital Twin

AI: Artificial Intelligence

AI4C2PS: Artificial Intelligence for Cognitive Cyber-Physical Systems Interoperability

AIGC: Artificial Intelligence-Generated Content

AIV: Autonomous Intelligent Vehicles

ANN: Artificial Neural Network

CDT: Cognitive Digital Twin

CI: Communication Interface

CPS: Cyber-Physical Systems

DAI: Data Acquisition Interface

DHM: Digital Human Modelling

DM: Digital Model

DS: Digital Shadow

DT: Digital Twin

DTS: Digital Twin System

EEG: Electroencephalography

EIF: European Interoperability Framework

EMMA: Eye Movements and Movement of Attention

FI: Feedback Interface

FINST: FINgers of INSTantiation

fMRI: Functional Magnetic Resonance Imaging

GAN: Generative Adversarial Network

GDPR: General Data Protection Regulation

GST: General Systems Theory

GSR: Galvanic Skin Response

HCC: Human-CPS Collaboration

HDT: Human Digital Twin

HDTS: Human Digital Twin System

HMI: Human-Machine Interface or Human-Machine Interaction

HODT: Human Operator Digital Twins

HRC: Human-Robot Collaboration

HRV: Heart Rate Variability

HUMO: HUman ONtology

ISO: International Organization for Standardization

JNH: Jumeau Numérique Humain

JNHa: Jumeau Numérique Humain augmenté

LIST: Luxembourg institute of Science and Technology

LLM: Large Language Model

LSTM: Long Short-Term Memory

MAE: Mean Absolute Error

MAS: Multi-Agent Systems

RAG: Retrieval-augmented generation

SEQ: Single Ease Question

SLR: Systematic Literature Review

SOSA: Sensor, Observation, Sample, and Actuator Ontology

SSN: Semantic Sensor Network Ontology

TAM: Technology Acceptance Model

UML: Unified Modeling Language

UTAUT: Unified Theory of Acceptance and Use of Technology

WDS: Workcell Digital Shadow

Résumé en Français (French Summary)

Contexte et questions de recherche

L'industrie évolue vers le paradigme de l'Industrie 5.0, qui privilégie une relation étroite entre l'humain et la machine. Pour instaurer cette relation, les machines doivent devenir conscientes de l'état physiologique, psychologique et cognitif des travailleurs humains. Les jumeaux numériques humain (JNH), représentations numériques synchronisées d'un individu permettant de suivre et prédire son état et son comportement, apparaissent comme une technologie candidate pour répondre à ce défi. Toutefois, l'adoption rapide du concept du JNH dans de nombreux domaines a entraîné une divergence des compréhensions conceptuelles, laissant peu de consensus sur ses aspects fondamentaux.

Cette thèse fait partie du projet AI4C2PS (ANR/FNR), qui vise à étendre l'interopérabilité entre humains et systèmes cyber-physiques (CPS) du niveau sémantique au niveau cognitif. Le projet suit une stratégie de jumelage numérique décentralisé où chaque humain et chaque CPS possède son propre jumeau numérique.

L'objectif général de la thèse est de poser les fondamentaux des JNH et d'étudier leur potentiel pour permettre aux CPS collaboratifs de s'adapter aux travailleurs humains. Trois questions de recherche structurent ce travail :

- **RQ1 :** Qu'est-ce qui définit un JNH et quels en sont ses éléments fondamentaux ?
- **RQ2 :** Comment les JNH peuvent-ils suivre et émuler la cognition et le comportement humains, et améliorer ainsi la collaboration homme-CPS ?
- **RQ3 :** Les architectures cognitives peuvent-elles modéliser la cognition humaine avec suffisamment de fidélité pour servir de modèle d'émulation aux JNH ?

Ces questions de recherche guident le travail, partant des aspects conceptuels vers l'application pour finir avec une validation expérimentale.

Organisation globale du manuscrit

Le manuscrit comprend une introduction, quatre chapitres et une conclusion, chacun répondant progressivement à ces questions de recherche :

- **Chapitre 1 :** revue systématique de la littérature sur les aspects conceptuels des JNHs, ses limites et le positionnement de la thèse.
- **Chapitre 2 :** proposition d'une définition générale, d'un système de classification et d'un méta-modèle des JNH ancré dans la théorie générale des systèmes.

- **Chapitre 3 :** proposition d'une architecture fondée sur les jumeaux numériques pour la collaboration homme-robot et formalisation de modèle d'émulation de la cognition humaine.

- **Chapitre 4 :** évaluation expérimentale de l'architecture cognitive ACT-R comme modèle d'émulation de la cognition humaine lors d'une tâche d'assemblage.

La conclusion fait synthèse des contributions au regard des questions de recherche, et discute des limites et perspectives.

Chapitre 1 - État de l'art

Ce chapitre présente une revue systématique de la littérature explorant les aspects conceptuels des JNH. La revue identifie un désaccord important sur le niveau d'intégration des données entre humain et son JNH, ainsi qu'un consensus sur l'entité jumelée qu'est un individu humain.

Le chapitre passe également en revue les défis majeurs du développement des JNH, directement liés à la nature humaine de l'entité jumelée : **(i) La collecte de données humaines** est limitée par rapport aux systèmes techniques : de nombreux aspects internes ne peuvent être captés que par des méthodes invasives, complétés par des données déclaratives ;**(ii) La modélisation** reste très spécialisée et centrée sur des cas d'usage étroits, loin de la vision ambitieuse d'un JNH persistant et holistique accompagnant l'individu tout au long de sa vie ;**(iii) Les aspects cognitifs** constituent un obstacle de modélisation majeur, la détection directe des états internes qui déterminent le comportement humain étant particulièrement difficile, contrairement aux machines.

À l'issue de cette revue, trois limitations structurent le positionnement de la thèse :

- **Limitation 1 : Ambiguïté sur les éléments fondamentaux des JNH.**

- **Limitation 2 : Absence de méta-modèle unifié pour le développement des JNH.**

- **Limitation 3 : Manque d'approches pour émuler les aspects cognitifs.**

Ces trois limitations définissent directement les objectifs de la thèse : ***(i)*** analyser les propriétés fondamentales des JNH pour proposer une définition générale et un système de classification indépendants du domaine ; ***(ii)*** développer un méta-modèle définissant les composants et interfaces du JNH ; ***(iii)*** formaliser les propriétés des modèles d'émulation de la cognition humaine et instancier une architecture cognitive comme modèle d'émulation pour une tâche d'assemblage.

Chapitre 2 - Fondamentaux du Jumeau Numérique Humain

Ce chapitre constitue la contribution conceptuelle centrale de la thèse : à partir des limitations identifiées au Chapitre 1, il propose une définition générale et un système de classification pour les JNH, ainsi qu'un méta-modèle ancré dans la théorie générale des systèmes. Ces deux contributions répondent conjointement à **RQ1** : *Qu'est-ce qui définit un JNH et quels en sont ses éléments fondamentaux ?*

2.1 Définition et classification

Le principal désaccord de la littérature concerne l'entité jumelée et le niveau d'intégration des données. Si le consensus définit l'entité jumelée comme un individu humain unique, deux minorités la voient comme une classe/rôle humain ou un système incluant des humains (e.g. un atelier). La thèse retient le consensus, préservant la relation univoque JNH-jumeau. Sur la question de l'intégration des donées la thèse propose une classification selon cet aspect suivant le concept de Digital Model, Digital Shadow et Digital Twin [Kritzinger et al., 2018]. En explorant un cas limit, la thèse introduit le concept d'*jumeau numérique humain augmenté* (JNHa) : un jumeau numérique jumelant un humain et un système technique fortement couplé (e. g. exosquelette), où les flux de données et changements d'état sont bidirectionnels entre les deux.

2.2 Méta-modèle systémique

S'appuyant sur la théorie générale des systèmes, le méta-modèle définit tout système par ses composants, son environnement, ses interfaces, ainsi que ses objectifs et fonctions. Il identifie quatre éléments essentiels au développement d'un JNH et leurs interfaces respectives, en distinguant systèmes physiques, systèmes numériques et systèmes cyber-physiques.

L'humain y est conceptualisé comme un agent qui répond aux stimuli de l'environnement par une action motrice : les organes sensoriels perçoivent le stimulus, le système cognitif prend une décision, puis l'action est exécutée. Le JNH reçoit des données sur l'état humain via deux canaux : des capteurs et des entrées manuelles converties en données exploitables par un dispositif d'interaction homme-machine. En retour, le JNH ne transmet pas directement des données à l'humain : sa rétroaction doit être transformée par le dispositif d'interaction en un stimulus physique perceptible. Le méta-modèle attribue au JNH trois fonctions majeures : (1) l'émulation, (2) les fonctions cognitives, et (3) la simulation.

2.3 Conclusion du chapitre

La définition générale et le système de classification répondent à la Limitation 1 (ambiguïté conceptuelle des JNH), tandis que le méta-modèle systémique répond à la Limitation 2 (absence de méta-modèle unifié). Ensemble, ces deux contributions répondent pleinement à **RQ1** et fournissent le cadre conceptuel exploité dans les chapitres suivants.

Chapitre 3 - Le Jumeaux Numérique Humain pour la Collaboration Humain-CPS

Ce chapitre restreint le champ d'investigation aux JNH dédiés à l'émulation de la cognition humaine et à leur potentiel pour permettre des adaptations dans des scénarios de collaboration homme-robot. Il répond partiellement à **RQ2** à travers deux contributions : une architecture fondée sur les jumeaux numériques et une formalisation de l'émulation cognitive.

3.1 Architecture fondée sur les jumeaux numériques

Le point de départ est un constat : de nombreux obstacles à la collaboration homme-robot et à l'interopérabilité cognitive découlent d'un manque de connaissance de l'état humain, en particulier de son état cognitif. Pour y répondre, le chapitre propose une architecture décentralisée où chaque entité collaborant possède son propre jumeau numérique : le travailleur humain, le robot et la cellule de travail elle-même.

- Le **Jumeau numérique humain** héberge un modèle d'émulation qui suit l'état cognitif du travailleur, y compris son intention, afin d'enrichir les informations disponibles pour les adaptations robotiques.

- Le **Jumeau numérique cognitif** du robot utilise des architectures cognitives pour sélectionner sa prochaine action de manière adaptative.

- Le **Workcell Digital Shadow (WDS)** centralise l'information sur l'état des ressources partagées et la progression de la production, la partageant avec les deux autres jumeaux.

Cette architecture est illustrée par deux cas d'usage : (1) une collaboration homme-robot visant à minimiser les conflits d'action, et (2) une tâche d'assemblage où l'humain reçoit du support personnalisé selon ses besoins spécifiques. L'approche décentralisée est conçue pour être extensible à un nombre arbitraire d'agents communiquant via leurs interfaces respectives.

3.2 Formalisation du modèle d'émulation cognitive

Parmi les défis identifiés pour cette architecture, la création d'un modèle d'émulation de la cognition humaine constitue l'obstacle le plus important. Le chapitre propose une formalisation mathématique d'un modèle de la cognition humaine, ainsi que deux algorithmes d'utilisation, en simulation et en émulation.

Le modèle d'émulation traite un stimulus et un état interne pour sélectionner une action et mettre à jour son état interne. En mode **émulation**, le modèle reste continuellement synchronisé avec les actions observées : à chaque étape, si l'action estimée diffère de l'action réellement observée, une fonction de synchronisation ajuste l'état interne du modèle sur la réalité observée. En mode **simulation**, le modèle est exécuté librement à partir d'un état initial jusqu'à un état cible, sans synchronisation, produisant une trace de comportement hypothétique.

Une exigence centrale de ce modèle est qu'il ne peut s'agir d'une simple boîte noire associant stimuli et actions motrices : l'interopérabilité cognitive requiert un modèle à état interne, capable de rendre compte des intentions et modèles mentaux du travailleur. Cette exigence motive le recours à une architecture cognitive plutôt qu'à des approches d'apprentissage automatique, car les architectures cognitives sont conçues pour reproduire les mécanismes fixes qui sous-tendent la performance humaine à travers différentes tâches, suivant le même cycle cognitif (perception, attention, mémoire, raisonnement) que leur jumeau physique.

Sur cette base, le chapitre explore le paysage des approches de modélisation cognitive et retient **ACT-R** comme architecture candidate, en raison de sa maturité et de sa validation empirique extensive sur plusieurs décennies. Le chapitre suivant présente la campagne expérimentale évaluant l'aptitude d'ACT-R à jouer ce rôle de modèle d'émulation lors d'une tâche d'assemblage.

Chapitre 4 - Validation Expérimental

Ce chapitre est consacré à la troisième et dernière question de recherche de la thèse (**RQ3**) : *Les architectures cognitives peuvent-elles modéliser la cognition humaine avec suffisamment de fidélité pour servir de modèle d'émulation aux JNH ?*. À travers une campagne expérimentale en deux volets, ce chapitre fournit une évaluation expérimentale de l'architecture ACT-R comme modèle d'émulation, en modélisant le comportement humain lors d'une tâche d'assemblage.

Expérience 1 : simulation de la tâche d'assemblage

L'objectif de cette première expérience est de tester la transférabilité d'ACT-R au domaine industriel, en évaluant sa capacité à simuler une séquence cohérente d'actions et à reproduire des tendances d'apprentissage et d'oubli semblables à celles d'un être humain, lors d'une tâche d'assemblage de composants stockés dans des tiroirs. Le modèle est exécuté en mode simulation pour produire des séquences d'actions.

Cette expérience établit avec succès la transférabilité d'ACT-R au domaine industriel : le modèle produit une séquence d'actions cohérente et reproduit des tendances d'apprentissage qualitativement plausibles. Elle révèle toutefois des limites importantes : les taux d'apprentissage du modèle et ses temps moteurs dépassaient largement les capacités humaines réelles.

Expérience 2 : émulation du comportement de recherche

Cette seconde expérience vise à évaluer si ACT-R peut émuler avec précision le comportement d'un individu au niveau de la réponse individuelle, en passant d'une simulation agrégée à une émulation personnalisée. Pour ce faire, le modèle intègre deux approches : la personnalisation pour refléter les différences individuelles, et la synchronisation du modèle à partir des réponses observées.

Les résultats montrent que la synchronisation améliore la précision du modèle (score F1) au niveau de la réponse individuelle, démontrant que l'alignement de l'état du modèle est techniquement réalisable, bien qu'un défi majeur subsiste : l'inférence des états cognitifs à partir des actions observées. En revanche, la personnalisation à l'aide d'un paramètre unique dérivé d'une tâche de calibration montre une faible transférabilité ; une personnalisation réussie nécessiterait d'encoder les stratégies mentales individuelles et les connaissances préalables de chaque participant.

Conclusions et lien avec les questions de recherche

Ces résultats apportent une réponse nuancée à **RQ3**. ACT-R peut être considéré comme une base prometteuse pour l'émulation de la cognition humaine, mais plusieurs limitations doivent être notées : le développement du modèle nécessite l'élaboration manuelle de règles de production spécifiques à la tâche, ce qui limite sa réutilisabilité et éloigne la vision d'un modèle général de cognition pour les JNH ; l'inférence des états cognitifs à partir de données comportementales reste une limite fondamentale, potentiellement restreinte à des tâches simples ; enfin, l'interaction d'ACT-R avec l'environnement physique demeure limitée.

Conclusions

Cette thèse a posé les fondations des JNH et exploré leur potentiel pour permettre aux systèmes cyber-physiques collaboratifs de s'adapter aux travailleurs humains. Les contributions apportées répondent aux trois questions de recherche comme suit :

- **RQ1** *(Qu'est-ce qui définit un JNH et quels en sont ses éléments fondamentaux ?)* est répondue par :
  - Une définition générale et un système de classification des JNH, indépendants du domaine, basés sur le niveau d'intégration des données.

  - Un méta-modèle ancré dans la théorie générale des systèmes, définissant les rôles, limites et interfaces essentiels des JNH.

- **RQ2** *(Comment les JNH peuvent-ils suivre et émuler la cognition et le comportement humains, et améliorer ainsi la collaboration homme-CPS ?)* est partiellement répondue par:
  - Une architecture fondée sur les jumeaux numériques pour la collaboration homme-robot, avec un jumeau numérique par entité (humain, robot, cellule de travail).
  - Une formalisation mathématique du modèle d'émulation de la cognition humaine et de son interaction avec l'environnement, accompagnée de deux algorithmes d'utilisation.

- **RQ3** *(Les architectures cognitives peuvent-elles modéliser la cognition humaine avec suffisamment de fidélité pour servir de modèle d'émulation aux JNH ?)* est partiellement répondue par :
  - Une évaluation expérimentale de l'architecture ACT-R comme modèle d'émulation, démontrant sa transférabilité au domaine industriel (Expérience 1) et les effets de la synchronisation et de la personnalisation sur la précision d'émulation individuelle (Expérience 2).

Dans l'ensemble, ces contributions établissent un cadre conceptuel cohérent pour les JNH, proposent une architecture concrète et un modèle formalisé pour la collaboration homme-CPS, et fournissent une première preuve expérimentale des possibilités et limites des architectures cognitives comme modèles d'émulation. Les limitations identifiées, notamment l'inférence des états cognitifs à partir du comportement observé et la faible transférabilité de la personnalisation, tracent des pistes claires pour les travaux futurs sur l'intégration d'un modèle d'émulation au sein d'un JNH pleinement fonctionnel.

Introduction

1 Context and Motivation

From mechanisation, electrification to digitalisation, industry has gone through major shifts leading to reimagining of the relationship between humans and technology. The Industry 4.0 paradigm follows a system-centric approach to push for highly inter-connected production systems with intelligent and autonomous local nodes [Oztemel and Gursev, 2020]. Cyber-physical systems are an enabling technology for Industry 4.0 being characterized by tight integration of the physical and cyber layer and their ability to communicate, compute and control [Baheti and Gill, 2011]. The long-term vision is to create fully automated, highly adaptive and resilient factories with minimal human interaction. This system-centric approach often implicitly relies on the myth of the "magic human", a design fallacy where operators are expected to handle all unforeseen system failures with perfect speed and coordination, despite being largely excluded from the control loop until a crisis occurs [Trentesaux and Millot, 2016]. However, Industry 5.0 pushes back and advocates for a closer relationship between humans and machines. This paradigm envisions the next generation of industrial systems along three key tenets; ***(i)*** resilience; ***(ii)*** sustainability; and ***(iii)*** human-centricity [Breque et al., 2021]. The shift from a system-centric to a human-centric view promotes going beyond creating safe and ergonomic environments for human workers, as it seeks to address higher needs like belonging, esteem and self-actualisation [Lu et al., 2022]. Humans are recognized as bringing valuable skills like creativity and decision making that are currently irreplaceable. A symbiotic relationship between human and machine is envisioned [Romero et al., 2016], such that human capabilities are enhanced, not replaced. For this close human-machine relationship, it is not enough for machines to react to humans' movements and actions. They need to be aware of humans' physiological, psychological and even cognitive state. Ultimately, they have to become "empathic machines" [Lu et al., 2022] and learn continuously to grow together with human co-workers. This human-centric transition builds upon the established role of Digital Human Modelling (DHM) in the digital factory, where populations of virtual manikins are utilised to analyse musculoskeletal loads and ergonomics during the early stages of development. This perspective aligns with current research trends which aim to enrich virtual models with mental features and embedded cognitive capabilities to bridge the gap toward true digital twins [Regazzoni and Rizzi, 2019].

The desire to have a closer human-machine relationship is in part a response to the pressure of mass customization where production runs tend to be short and product variety high. Human workers are highly flexible and creative making them indispensable to adapt to the high variety of product types and production paces [Ivanov, 2023]. Nonetheless engineered systems, like robots, come with important strength on their own like their reliability, accuracy and high repeatability. The field of human-robot collaboration seeks to combine the strength of both robots and humans by creating collaborative setups in close physical proximity. But the presence of a robot can easily lead to increased stress levels and represent an injury risk [Wang et al., 2019]. As humans we

have the ability to understand each other's intent and thought processes through gestures, gazes and of course open communication. Robots can be equipped with sensors to detect the physical movement and behaviour of collaborating humans but are mostly blind to their internal state such as intentions. This blindness can lead to misunderstandings, poor coordination, degraded performance, collision risk, and ultimately an erosion of trust. To ensure good collaboration and enable robots to be active collaborators they have to move beyond reacting to physical movement. They require an awareness of human intent to dynamically adapt to their behaviour in a shared workspace.

Human Digital Twins (HDT) represent an ideal candidate technology to address this challenge in human-robot collaboration or human-CPS interaction more widely. An HDT is a synchronised digital representation of a human individual, tracking their state and making predictions of future states and behaviour possible. They centralise personal data about their human twin and can selectively share information with collaborating system, enabling them to become human-aware and adaptive. The HDT concept first appeared in industry [Graessler and Poehler, 2017] and since spread to numerous new domains, but this rapid spread led to a divergence in conceptual understanding. Therefore, agreement on fundamental aspects that constitute an HDT is poor and requires clarification. Traditional Digital Twins (DT) of engineered systems are characterized by automatic bi-directional communication between the physical entity and DT [Kritzinger et al., 2018]. In HDT research there is no clear agreement on whether a bi-directional communication link is a necessary component of HDTs [Lauer-Schmaltz et al., 2024]. From this lack of agreement, theoretical and technological advancements risk being incompatible or contradictory hampering the progress of wider HDT research.

A wide variety of challenges for HDT development are a direct consequence of the twinned entity's human nature. The inherently personal nature of the gathered data poses ethical concerns while gathering high quality continuously is a technical challenge in its own. First and foremost, humans are highly complex natural systems that are only partially understood and cannot easily be represented by computerized models. For instance, human behaviour is the result of not only cognitive and physiological processes, but also personality traits and personal biases. In a human-robot collaboration context, information about human intention would enable robots to adapt and avoid conflicting actions but capturing human intent or estimating it is a major obstacle. Direct detection of cognitive states faces major challenges. It is limited and relies on cumbersome devices that limit mobility of the worker such as those implementing EEGs and fMRIs.

A less invasive alternative are *cognitive architectures*. They require information about the stimuli the human is sensing, which can be detected with less invasive methods like eye-tracking. They are computational models that follow the known structure of the human brain [Newell, 1990; Laird et al., 2017]. A large variety of cognitive architectures exists [Kotseruba and Tsotsos, 2020] and are designed to either give cognitive ability to engineered systems or model the cognitive processes of humans [Al Haj Ali et al., 2024a]. In the latter category, ACT-R is the most mature model. Since the start of its development in the 1976 (under the name ACT), it has been applied in numerous contexts ranging from psychological experiments over linguistics to Human-Computation interaction [Kotseruba and Tsotsos, 2020]. It is specialized in modelling human behaviour and successfully reproduces numerous experimental results from cognitive science and neuroscience [Anderson et al., 2004]. Based on the internal state and received stimuli, it simulates the cognitive process including goals and action selection. The model has succeeded in simulating human behaviour and performances when performing various tasks [Ritter et al., 2019]. However, ACT-R stops short of emulating the behaviour at the granularity of the individ-

ual action. Emulation requires the model to reflect an individual's cognitive abilities to account for important interpersonal differences. A second challenge is to keep the model synchronized with the state of the twinned human.

This thesis sets itself the objective to clarify the fundamentals of HDTs, particularly for industrial applications involving Human-CPS collaboration (HCC). As a second objective, the thesis seeks to formalise emulation of human cognition and investigates the potential of cognitive architectures to function as an emulation model within an HDT.

1.1 Project context: AI4C2PS

This thesis is conducted within the framework of the *Artificial Intelligence for Cognitive Cyber-Physical Systems Interoperability* (AI4C2PS) project, a bilateral initiative funded by the French National Research Agency (ANR) and the Luxembourg National Research Fund (FNR). The consortium includes the University of Lorraine, the Research Center for Automatic Control (CRAN CNRS UMR 7039), the Luxembourg Institute of Science and Technology (LIST), and the private company Orisun (ORSN)

The project focuses on establishing continuous interoperability within collaborative enterprise systems where Cyber-Physical Systems (CPS) and humans work together toward common objectives. It addresses a critical limitation in current industrial systems: while traditional ontology-based approaches ensure semantic interoperability between technical components, they are insufficient for enabling mutual understanding between humans and machines, as these entities process knowledge in fundamentally different ways. To bridge this gap, AI4C2PS proposes extending interoperability from the semantic to the cognitive level by integrating models that represent concepts, cognitive processes and collaborative dynamics. The project's strategy relies on a decentralized digital twinning, where every CPS and human individual is paired with its own Digital Twin. This decentralized approach informs the thesis's proposal to move away from aggregate shop-floor models in favour of personalized, persistent HDTs that can accompany an individual across various contexts. Within AI4C2PS, the HDT is positioned as a key technology to emulate human cognitive processes at runtime to assure alignment of understanding between collaborating entities and contribute to cognitive interoperability.

2 Research Questions and Contributions

The overarching objective of this thesis is to *lay the foundations of Human Digital Twins and investigate their potential to enable collaborating CPS to adapt to human workers in industrial contexts.* To achieve this objective multiple separate research questions are defined:

RQ1: *What defines a Human Digital Twin and what are its fundamental elements?*
The relative novelty of HDTs and their rapid spread leaves room for ambiguity in its definition and core aspects. This research question aims at building a solid foundation for HDT developments by ***(C1)*** proposing a clear general definition and developing a classification system; ***(C2)*** building a meta-model to clarify the essential components of an HDT.

RQ2: *How can HDTs be used to track and emulate human cognition and behaviour? In particular, how can this improve collaboration between humans and Cyber-Physical Systems?*

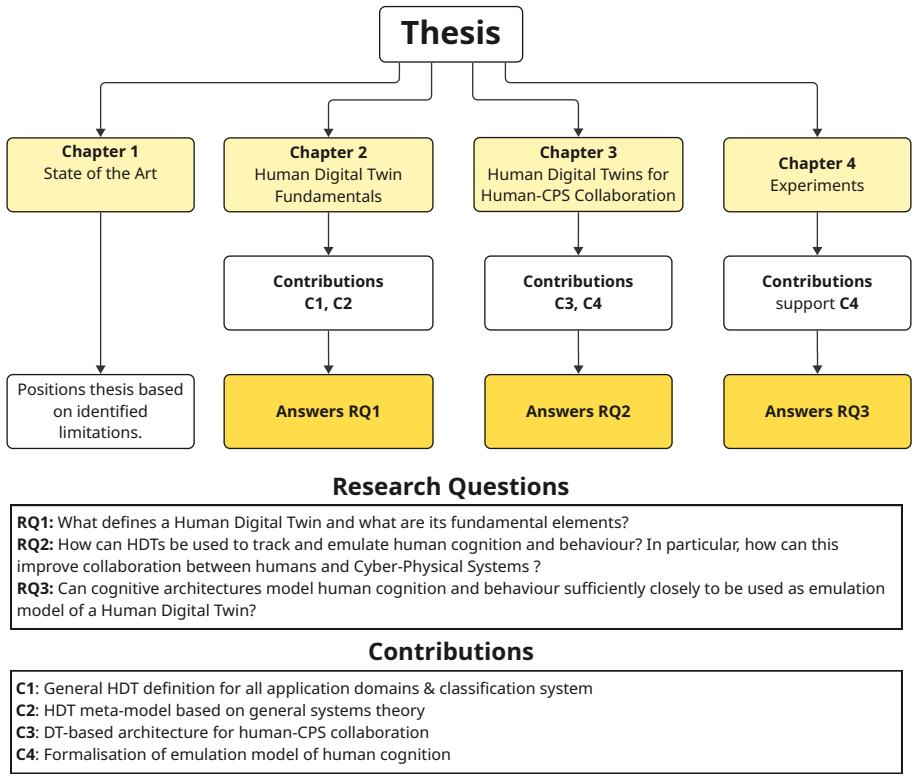


Figure 1: Structure of the thesis and contributions for each chapter.

An HDT is closely integrated with its twinned human. This close relationship makes it a promising candidate to receive and process behavioural data that may prove useful in Human-CPS Collaboration (HCC). With this research question, we aim at ***(C3)*** clarifying how gathering, processing, and sharing behavioural data from a twinned human through an HDT can improve HCC, and ***(C4)*** formalising an emulation model of human cognition.

RQ3: *Can cognitive architectures model human cognition and behaviour sufficiently closely to be used as an emulation model for Human Digital Twins?*
Human cognition is notoriously hard to model. While cognitive architectures exist to model it based on the human brain structure, it is currently not used for emulation purposes. With this question, we seek to ***(C4)*** explore the landscape of cognitive architectures for emulation purposes and experimentally evaluate whether the selected architecture is able to emulate an individual's behaviour on the level of the individual action.

3 Structure

Figure 1 provides an overview of the thesis structure. The first two chapters focus on the literature and conceptual foundations of Human Digital Twins in general. The last chapters reduce the scope to Human-Robot Collaboration use cases and the challenge of modelling human cognition.

Chapter 1 - State of the Art: This chapter presents a systematic literature review exploring the conceptual foundation of Human Digital Twins [Gaffinet et al., 2025a]. A detailed description

of the evolution of the concept is presented, including the spread to new application domains and emergence of conflicting views on foundational aspects. The review identifies contemporary challenges for HDT development and positions the contributions and objectives of this thesis. *This chapter lays the foundation of the thesis and provides critical information for RQ1.*

Chapter 2 - Human Digital Twin Fundamentals: Based on the conflicting views on HDT fundamentals, this chapter analyses the nature of the divergences and addresses them [Gaffinet et al., 2025a]. Fundamental differences between HDTs and traditional DTs are discussed to extract the requirements that define HDTs. These requirements are synthesized to propose: ***(i)*** a general definition for any application domain and classification based on level of data integration *(contribution C1)*; ***(ii)*** a meta-model anchored in general systems theory *(contribution C2)* [Gaffinet et al., 2025c]. *This chapter answers RQ1 through contribution C1 and C2.*

Chapter 3 - Human Digital Twins for Human-CPS Collaboration: This chapter constrains the scope of the investigation to HDTs for the emulation of human behaviour and how it can enable adaptations in HRC setups. First, we discuss how an HDT emulating human behaviour can be exploited to enable human-aware engineered systems. We propose a DT-based architecture for human-robot collaboration based on the systemic HDT meta-model [Gaffinet et al., 2025b, 2023] *(contribution C3)*. Second, we formalise the emulation model for human cognition that is hosted by the HDT *(contribution C4)* and investigate modelling approaches to implement emulation models. *This chapter answers RQ2 through contribution C3 and C4.*

Chapter 4 - Experimental Contributions: Previous chapters focus on conceptually clarifying how HDTs and cognitive emulation models may serve as key enablers for adaptive HRC. This chapter presents two experiments performed in laboratory conditions. A cognitive architecture, ACT-R, is instantiated to emulate observed human behaviour during an assembly task. Models are personalised to reflect the cognitive abilities of individual participants and synchronised with observed actions. This setup allows evaluating ACT-R's suitability for emulation purposes and quantify the impact of personalisation and synchronisation. The chapter ends with a discussion on limitations and future research directions. *This chapter provides supporting evidence for RQ3 and supports contribution C4.*

Chapter 5 - Conclusions: This final chapter summarises the contributions of this thesis and how they contribute towards answering the stated research questions. A discussion on limitations and future research directions closes this thesis with the hope to provide guidance for further research on Human Digital Twins and emulation of human cognition.

Chapter 1

State of the Art

1.1 Introduction

The concept of the Digital Twin (DT) was introduced by NASA during the early Apollo program [Allen, 2021]. It was further developed as a software and applied to industry in the 80s and 90s[Corbier, 1989; Panetto, 1991; Iung et al., 2000] before being widely popularized by Grieves [2002]. Within the manufacturing context, a DT is defined as a *"fit for purpose digital representation of an observable manufacturing element with synchronisation between the element and its digital representation"* [ISO-23247, 2020]. Typical application cases are tracking the status of a machine and triggering maintenance cycles predictably before expensive damage occurs, or triggering an emergency stop if parameters drift far from their nominal values. Aligned with the tenants of Industry 5.0, a new type of DT, known as Human-Centric Digital Twin was proposed [Kider, 2025]. It includes the state of the human worker with the objective to enable systems that adapt and respond to human needs. An illustrative example is a DT that monitors worker fatigue to dynamically reallocate tasks. However, the Human-Centric Digital Twin modelling scope is a system containing human participants and stops short of creating a holistic model of the humans themselves. An additional ethical concern is the collection of personal data to be used in a larger system. In contrast, *Human Digital Twins (HDT)* focus specifically on digitally representing humans, rather than treating them as just another element of a larger system, making them well-suited to handle and protect personal data from their human counterparts. In the long-term vision, an individual would be accompanied by their HDT throughout life to provide personalised assistance in a large variety of contexts and situations. The HDT would host interlinked models about numerous aspects of human nature (e.g. physiology or cognition), learn about individual preferences and evolve over time through continuous learning.

In practice, creating a DT of a human comes with a broad variety of specific challenges. Machines can be designed with integrated sensors to provide exploitable data to their DTs, which is only true in a limited fashion for humans. For instance, humans understand their environment through stimuli which are hard to capture and feed to the HDT, not to mention the difficulty of modelling the subsequent cognitive processes transforming the stimuli into a specific reaction. Methods to detect internal human processes are limited and often invasive, which makes any modelling attempts significantly harder than for machines. A large body of literature addresses the technical challenges of gathering human data and building models on top of it.

HDTs have first emerged in Industry where traditional DTs are well understood. More recently HDTs spread to new communities, bringing not only new perspectives and interpretations, but also confusions about the concept. A key aspect of traditional DTs is the high level of data

integration, i.e., the data flow between the DT and the physical entity. However, in Healthcare we see models without persistent data flows that are identified as HDTs, for instance HDTs that are used as simulation platforms [Sinisi et al., 2020].

Just as it was important to distinguish between digital models and digital twins for traditional DTs in industry [Kritzinger et al., 2018], it is essential to clarify the vocabulary for HDTs. Building a human model is already a considerable challenge, requiring interdisciplinary efforts from fields like cognitive science, psychology and biomedicine. But synchronizing a model with real-world changes adds further complexity requiring expertise from, among others, computer science and communication systems communities. To facilitate the required collaboration and identify the right challenges, it is imperative to clarify the conceptual basis for HDTs. Specifically in industry, it is important to determine whether HDT refers to a DT of an industrial system that considers human interactions and the influence of human factors, or if it refers to a DT of the human entities within the system.

During this thesis we performed a systematic literature review to extract and analyse their interpretation of what constitutes an HDT and identify current challenges to position the work of the thesis. The full methodology and results from the review were published in Gaffinet et al. [2025a].

1.2 Prior Reviews

This section discusses prior reviews and their contents to identify the research gap we seek to address with the systematic literature review. The prior reviews come from authors with various backgrounds which informs their point of view and understanding and restricts the scope of their efforts. From Industry multiple reviews are investigating topics that are closely related and overlapping with HDTs, namely the topics of Human-centric DTs [Asad et al., 2023] and Human-Robot Collaboration DTs [Ramasubramanian et al., 2022]. One review from Industry is exclusively focused on HDTs [Wang et al., 2024a]. From Healthcare four reviews that exclusively focus on HDTs have been published [Sahal et al., 2022; Chen et al., 2024c; Lauer-Schmaltz et al., 2022; Chen et al., 2024a]. From the field of Mobility, a pivotal review in Hu et al. [2022] has provided a solid basis for HDT developments in their discipline. Finally, a set of two reviews are general in scope [Miller and Spatz, 2022a; Lin et al., 2022] without restricting themselves to a specific application domain. Besides the application area, the reviews also differ in their focus. Some are interested in the conceptualisation of HDTs while others explore the enabling technologies to provide guidance for designers and developers. See Table 1.1 for an overview on the domain, scope and focus of each review. We elaborate on the content of the past reviews and the perceived research gap we seek to address.

First parts of the HDT concept emerged in Healthcare , with the generation of patient specific personalised models for surgical planning[Passera et al., 2013]. The first appearance of the full HDT concept is from Industry, with Graessler and Poehler [2017]. In 2022, Lauer-Schmaltz et al. [2022] performed a review on HDTs in Healthcare for behaviour change therapy. Despite the highly specific use case the paper discusses fundamental differences between traditional DTs and HDTs. At the center is the personal nature of medical data and an increased need for stringent data safety and privacy. Ethical questions, for instance whether an HDT should be deleted after death, are raised. To maintain trust in the HDT all decisions need to be explainable, and models based on human data need to deal with poor data quality and common data gaps.

Reference	Domain	Scope	Focus
Asad et al. [2023]	Industry	Other DT	Pragmatic
Chen et al. [2024c]	Healthcare	HDT	Pragmatic
Chen et al. [2024a]	Healthcare	HDT	Pragmatic
Hu et al. [2022]	Mobility	HDT	Pragmatic
Lauer-Schmaltz et al. [2022]	Healthcare	HDT	Conceptual
Lin et al. [2022]	General	HDT	Conceptual
Miller and Spatz [2022a]	General	HDTS	Conceptual
Ramasubramanian et al. [2022]	Industry	Other DT	Pragmatic
Sahal et al. [2022]	Healthcare	HDT	Pragmatic
Wang et al. [2024a]	Industry	HDT	Pragmatic
Gaffinet et al. [2025a]	**General**	**HDT**	**Conceptual**

Table 1.1: Classification of prior reviews and positioning of our own review.

Capturing data about human cognition and modelling it is a unique challenge stemming from the human nature of the twinned entity. The paper provides a set of design considerations for how to develop an HDT, but clearly states that the definition, and essential components, of the HDT concept are underdeveloped and require further research.

Two reviews focus on conceptual aspects of HDTs without a specific application case [Miller and Spatz, 2022a; Lin et al., 2022]. In Miller and Spatz [2022a] the authors first propose a Digital Twin System (DTS) which is composed of a Real-World Twin, the DT itself, and an Interchange Component. As an extension a Human Digital Twin System (HDTS) follows the same structure but contains at least one human agent that is modelled alongside other machinery, sensors and actuators. This digital representation is considered an HDT, but it can be created, not only for a human individual but also for a human class. The authors take the position that the HDT can take the shape of a purely mathematical model and does not require a direct link with a physical human. The differentiation between HDT and HDTS represents a significantly different conceptual point of view.

A subsequent review by Lin et al. [2022] proposes a more detailed HDTS framework. However, their framework exclusively focuses on a human individual and their digital model, without any surrounding entities. In their view an HDT conceptually requires a twinned human individual, real-time data gathering and modelling of the human. The authors further elaborate on the limitations and challenges of creating HDTS. A major limitation lies in the gathering and transmitting of heterogeneous data, while guaranteeing its safety and reasonable latency. The authors underline the need for ontologies that unify human body data. Societal challenges include ethical questions as well as problems related to trust and social regulation.

A second set of reviews has an increased focus on specific application domains. They are less concerned with the conceptual basis for HDTs and focus their efforts on enabling technologies, frameworks and architectures to enable future HDT developments. For the domain of Mobility applications an influential review by Hu et al. [2022] applies the concept of HDTs to drivers. The review proposes a detailed framework to build a Driver Digital Twin with the goal to detect driver attention, distractions, intentions and drowsiness. For each of the four applications guidance is provided including taxonomies of the possible types of distractions and driver states. While the application domain is highly specific many identified challenges are general in nature. Both heterogeneous data and highly specialised scattered models are challenges. To develop the most

impactful Driver Digital Twin holistic models are needed, but multi-model datasets are lacking. Another general challenge is the need for personalisation, especially when modelling behaviour which is highly dependent on the individual's characteristics. The authors distinguish the Driver DT from the Driver DT System, the former being a digital representation of a driver HDT, while the latter includes interconnected components from the car.

Within Healthcare the most investigated topic in terms of performed reviews is personalised Healthcare. An extensive analysis of Personal Digital Twins is presented in Sahal et al. [2022], which can be seen as synonymous with HDT. They discuss the benefits, potential applications, essential building blocks and open challenges. Particular attention is given to artificial intelligence and blockchain technology and how they can be integrated for enhanced personalised Healthcare services. A second extensive review by Chen et al. [2024c] investigates many facets of the concept for personalised medicine. One observation is how separate DTs for individual organs can merge into one holistic HDT for a patient. The central topic of the review is a network architecture that is divided in 5 layers; data acquisition; data communication; computation; data management; data analysis and decision making. For each layer the authors address in detail the enabling technologies and spend little time on conceptual questions. A third review with an interest in personalised Healthcare is presented in Chen et al. [2024a], where the authors perform an extensive review on how generative artificial intelligence techniques may address key challenges in HDT development. In particular the potential to generate synthetic data, enhance communication and data management, creating better digital models and performing data analysis is investigated thoroughly. All three reviews view an HDT as a digital replica or representation of a human being that evolves with time as the human changes. Two of them state that real-time data is needed [Chen et al., 2024c,a], implying continuous and automatic data acquisition, while the third review is not explicit on this matter [Sahal et al., 2022].

Wang et al. [2024a] perform a review of HDTs for industrial cases, and specifically Industry 5.0. The central contribution is a system architecture organised in four layers to guide future developments. For each layer a detailed discussion of enabling technologies is provided. The conceptual grounding is not a central point in the paper, but the authors align with the view that an HDT is necessarily twinned to a human in the physical world, with bi-directional data flow.

Ramasubramanian et al. [2022] review DTs for Human-Robot Collaboration and identifies approaches related to HDTs. Operator data is included in the DT, but the modelling efforts are mostly limited to the creation of 3D avatars and human motion. The review discusses the challenges of Human-Robot Collaboration which notably includes the dynamic nature of the human and a need to account for their behaviour. Modelling a human operator is recognized as extremely complex and difficult. Current technology is seen as insufficient to create a holistic Operator Digital Twin, that includes not only physical, but also emotional aspects.

Most recent reviews focus on enabling technologies and provide guidance to developers through guidelines, frameworks and architectures. From the three reviews that address conceptual aspects; one calls for further research into essential components and the definition [Lauer-Schmaltz et al., 2022]; one sees HDTs as digital representations of a human [Lin et al., 2022]; while the last proposes an alternative view where HDT may also represent a human class [Miller and Spatz, 2022a].

Our published review [Gaffinet et al., 2025a] addressed this gap in two steps; ***(i)*** describing the evolution of the HDT concept and the diversity of conceptualisations; ***(ii)*** analysing these differences to propose a general definition and categorisation system. The remainder of this chapter presents the evolution of the HDT concept alongside the main identified challenges to

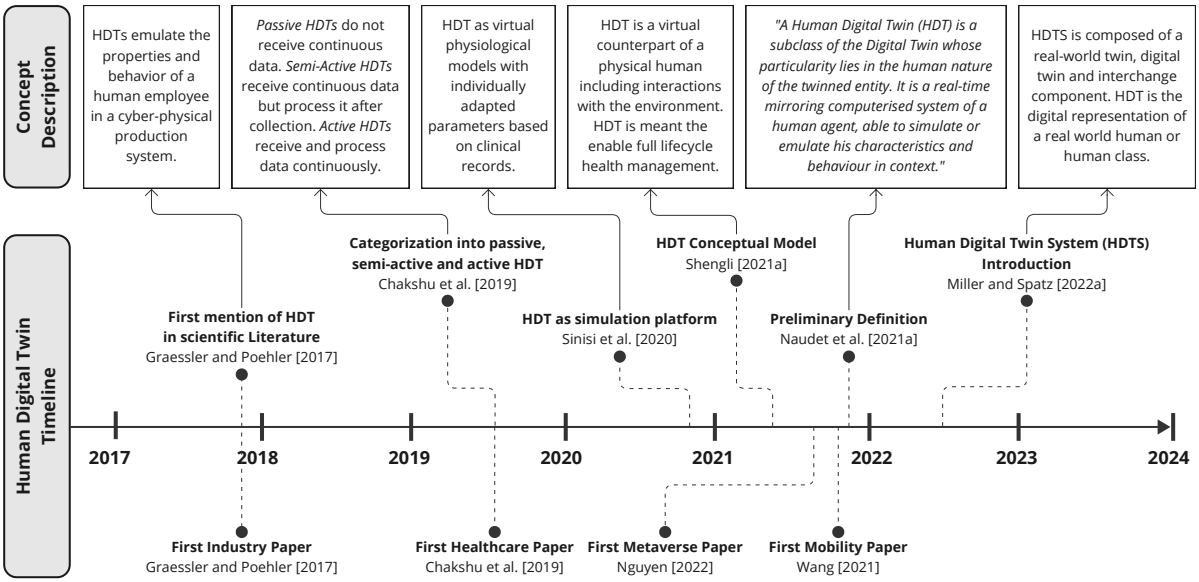


Figure 1.1: Timeline of Human Digital Twins (HDT) and various conceptual perspectives.

position the thesis and its contributions.

1.3 Descriptive Analysis

The objective of the literature review is to describe and analyse the different viewpoints when it comes to the definition of an HDT. A way of categorising the papers is by application areas. The concept of HDT was born in the community of industrial research, where the concept of DTs itself has strong roots. Over time the concept has spread to other application domains such as Healthcare, Metaverse applications or Mobility. Both the variety of applications and ways of conceptualising the HDT have grown over time. In this part of the chapter, we present the main application areas and how the number of publications has evolved over time in Section 1.3.1, a quick look at the geographic distribution of research in Section 1.3.2, and a detailed exploration into the various conceptual understanding in Section 1.3.3.

1.3.1 Timeline & Application Areas

This section presents the evolution of the HDT concept and application areas. Figure 1.1 shows the key moments in HDT research, namely the appearance of new conceptual perspectives, and entry into new major application areas.

We have gathered all papers on HDTs, including papers from all application domains of the papers. Four main domains emerged from our selection: ***(i)*** Industry (32.2%), ***(ii)*** Healthcare (28.9%), ***(iii)*** Metaverse (7.8%) and ***(iv)*** Mobility (5.5%). The first scientific publication that uses the concept of HDT was published in 2017 with an application in industrial production systems [Graessler and Poehler, 2017]. The author applies the concept of DT to a human to capture their behaviour and act as their representative. Since the first publication the concept has attracted increased interest, which manifests as an accelerated growth in the number of publications, as presented in Figure 1.1. Notably Industry was the first application area to

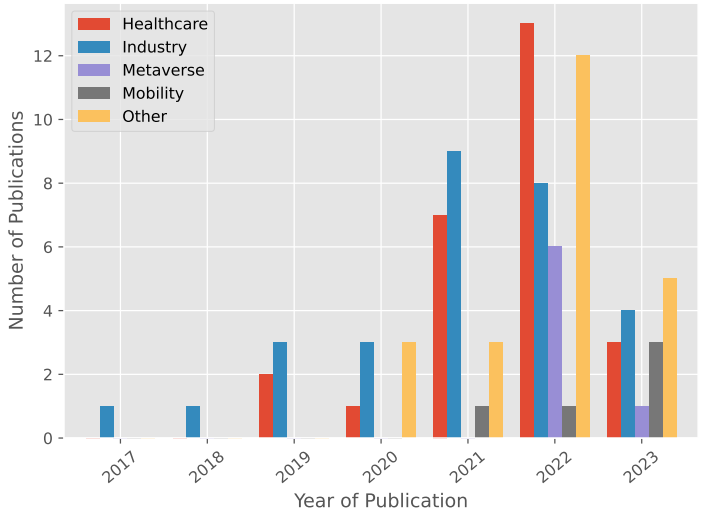


Figure 1.2: Number of publications by year and by application area. In 2023 papers were screened until the 15th of June.

adopt the concept, and an increased number of publications follows over the next years. In 2019 the concept spreads for the first time to another domain, namely into Healthcare [Chakshu et al., 2019]. Within only 3 years the number of publications in Healthcare surpassed the ones from Industry and is now the most active application area in terms of number of scientific publications. Since 2020 the concept of HDT is entering many different application areas. The number of application areas has been increasing since. Two notable application domains that emerged later are Mobility, and Metaverse applications. Research in Mobility applications started in 2021 [Wang, 2021] and has experienced a slight increase since.

The total number of scientific publications by application area is shown in Figure 1.3(a). HDT research remains dominated by Healthcare and Industry with a combined 61.1% of all extracted scientific publications. Both Metaverse and Mobility applications are new additions that are worth setting apart. Notably a large number of publications is categorised in the *Other* application area category. This is due to the rapid introduction of HDTs in completely new, and highly diverse area. From Figure 1.2 the number of publications that do not fit prior categories is increasing rapidly around 2022. A more detailed breakdown on Figure 1.3(b) illustrates how wide the application areas have spread since 2022.

1.3.2 Geographic Distribution

The start and evolution of HDT research divided by continent are shown in Figure 1.4. First publications start in Europe with a very low volume of publication for 3 years, before growing rapidly. By 2021 the research groups from Africa, North America and Asia have all published research on HDTs. A more detailed breakdown of publications by country and application area is presented in Figure 1.5. Notably the United States of America is leading in terms of total number of scientific publications and is the only country to publish for Mobility applications.

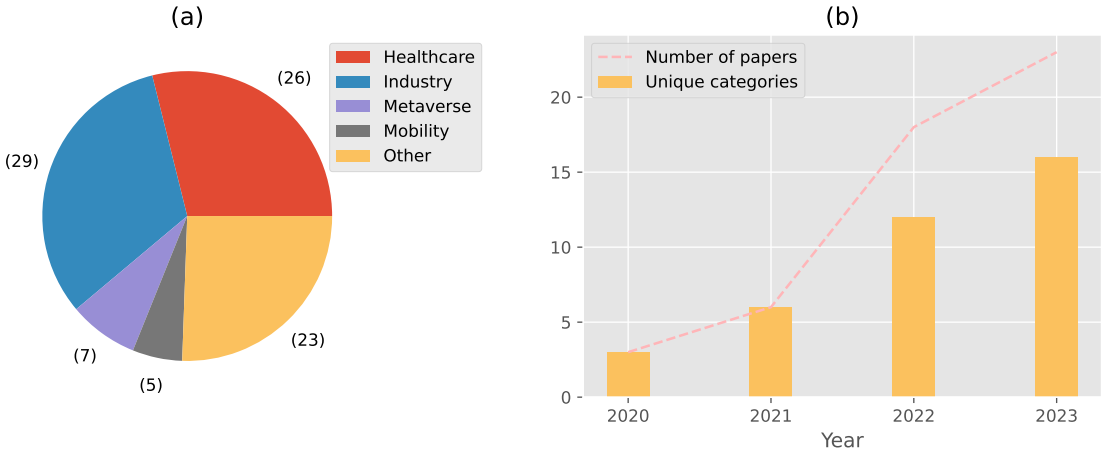


Figure 1.3: (a) Total number of publications by application area. (b) Total number of papers categorised as "Other" and number of unique categories that have less than 4 papers.

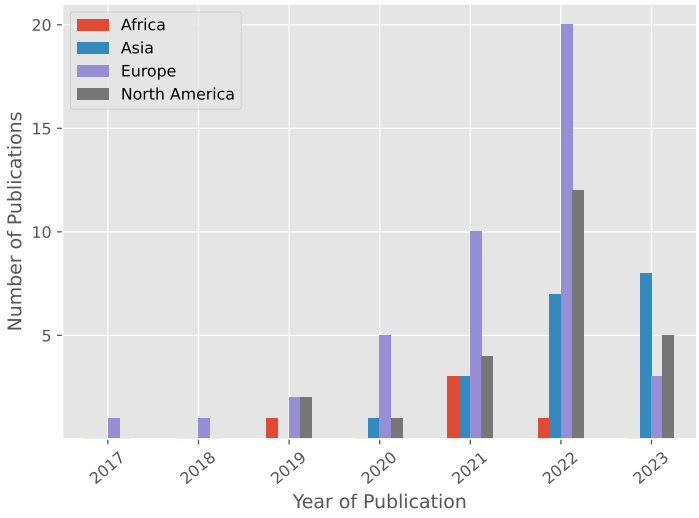


Figure 1.4: Number of publications by year and continent.

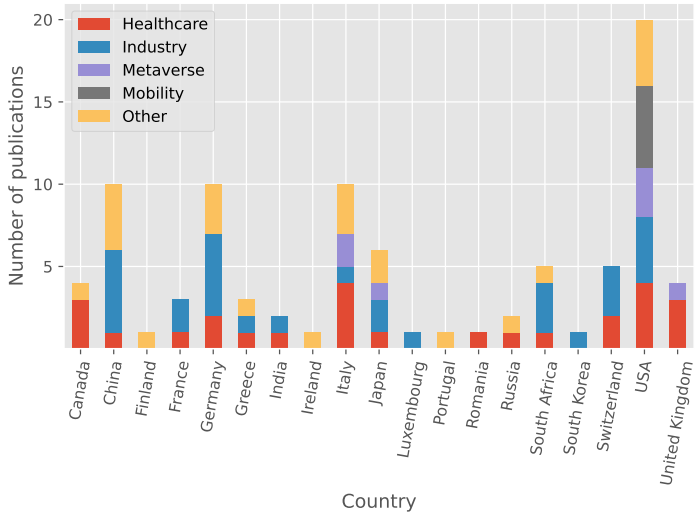


Figure 1.5: Number of publications by country and application area.

1.3.3 Concept of HDT

Due to the high number of different application cases many different scientific communities are researching HDTs. With each community comes a different set of domain-specific vocabulary and understandings of what an HDT should consist of. In Figure 1.1 we highlight the appearance of conflicting perspectives on the HDT concept. A point of disagreement is the nature of the data flows between the human and HDT. To be precise, HDTs are presented as simulation models without continuous synchronisation [Sinisi et al., 2020; Chude-Okonkwo, 2021] or as a fully synchronised digital representation that maintains bi-directional data flow with its human [Shengli, 2021a]. A second point of contention is the scope of the HDT's real-world twin. In Miller and Spatz [2022a] an HDT may model a class of humans, while a more common conception is that HDTs model human individuals.

The objective of this literature review is to bridge the gap among scientific communities and provide a general, domain-agnostic definition for HDT. Prior work proposed a preliminary definition which serves as starting point for the present systematic literature review:

A Human Digital Twin (HDT) is a subclass of the Digital Twin whose particularity lies in the human nature of the twinned entity. It is a real-time mirroring computerised system of a human agent, able to simulate or emulate his characteristics and behaviour in context. [Naudet et al., 2021a]

We described the current landscape of HDT definitions and conceptualisations across application domains. An important number of papers do not introduce or define the concept of HDT before using it. Indeed 38.3% of papers don't reference any existing definition nor do they introduce the concept explicitly. Because of this shortcoming we need to extract the implicit understanding of the HDT concept from their implementation or proposed usage. From the preliminary definition in Naudet et al. [2021a], a few key dimensions can be extracted.

1. **Digital Twin:** The concept of DT first found numerous applications in Industry. However, it is a new concept for the Healthcare community. The perception of what constitutes a

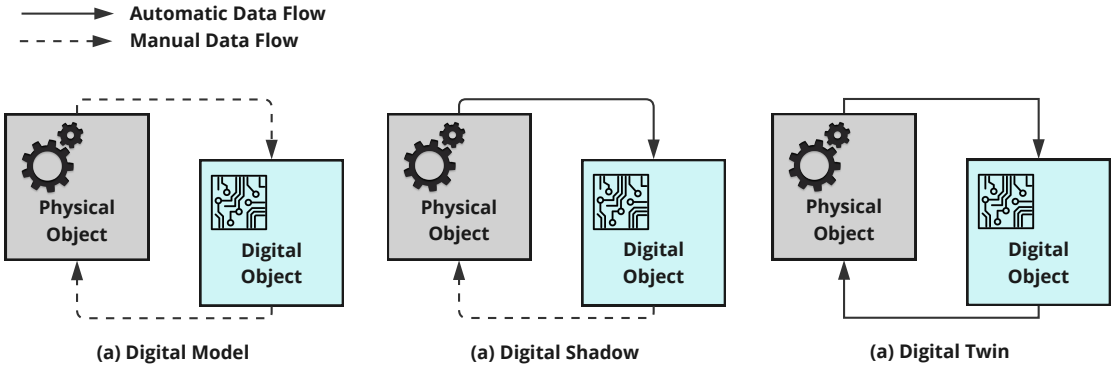


Figure 1.6: **(a)** Digital Model, **(b)** Digital Shadow and **(c)** Digital Twin as introduced in Kritzinger et al. [2018]. The categories are distinguished by the level of data integration.

DT is not uniform within or across scientific communities. We discuss the definition and types of DTs encountered in literature when applied to humans in Section 1.3.3.

2. **Twinned Entity:** The twinned entity of an HDT needs to be of human nature. From the scientific literature we extracted what the twinned entity of HDTs is and present it in Section 1.3.3.

Digital Twin Conceptualisation

A clear understanding of the DT concept is a fundamental part of any HDT definition. At its core, a DT is at least composed of a twinned physical system and a digital object that contains all measurable information about the physical system [Grieves and Vickers, 2017]. A later popular definition and model includes 5 dimensions, [Tao et al., 2018; Qi et al., 2021]; (1) physical entity (2) DT Data (3) Virtual Models (4) Services (5) Connections between each of the dimensions. Further extensions for specifically Human-Machine DTs have been proposed in Ascone and Vanderhaegen [2022], which explicitly include the environment, the human user, and a discussion on shared autonomy between the human and their DT.

In 2018, a major paper by Kritzinger et al. [2018] presents how the concept of DT is used differently in various disciplines, as well as conceptual variations in the definitions themselves. According to the authors the level of data integration of the DT is not always considered explicitly. To harmonise the understanding they introduce three subcategories which are distinguished by their level of integration (see Figure 1.6), namely; ***(i)*** Digital Model, ***(ii)*** Digital Shadow, and ***(iii)*** Digital Twin.

1. ***(i)*** **Digital Models** represent physical objects without any automated data exchange (e.g. simulations): *"A change in state of the physical object has **no** direct effect on the digital object and vice versa."* [Kritzinger et al., 2018]

2. ***(ii)*** **Digital Shadows** receive data from the physical object through an automated data flow: *"A change in state of the physical object leads to a change of state in the digital object, but not vice versa."* [Kritzinger et al., 2018]

3. ***(iii)*** **Digital Twins** have integrated and automatic data flows in both directions, which enables the DT to act as a control instance: *"A change in state of the physical object directly leads to a change in state of the digital object and vice versa"* [Kritzinger et al., 2018].

As an additional anchor for our understanding of DTs, we use a more recent definition as provided; *A set of adaptive models that emulate the behaviour of a physical system in a virtual system getting real time data to update itself along its life cycle [...]* [Semeraro et al., 2021]. For our own analysis we extract three pieces of information from the scientific papers:

1. Have the authors explained the concept of DT before using it? Do they reference an existing DT definition, explicitly proposing a definition or explaining the concept in their own words all count?

2. We distinguish between papers that propose conceptual models, frameworks or architectures for HDTs (*concept* papers) from papers that describe implementations (*case study* papers). Authors that propose a conceptual approach and implement it in the same paper are counted as *case study* papers only.

3. Finally, we categorise the conceptual models or implementations according to their level of integration following the three sub-categories from Kritzinger et al. [2018] as shown on Figure 1.6.

Table 1.2: Number of publications grouped by level of integration and whether they provide a definition before using the HDT concept.

Definition	**Digital Twin**	**Digital Shadow**	**Digital Model**
Yes	39	5	12
No	10	6	18

Out of the 90 papers on HDT analysed, 49 were about true DTs, 11 about Digital Shadows and the remaining 30 about Digital Models. The level of integration is changing significantly depending on whether authors provide a definition or not in their paper (See Table 1.2). Papers that don't provide a definition tend to follow the tenants of DTs less closely and have increased risk of developing Digital Models instead. It indicates a potential misidentification of the concept.

A second major factor that impacts the integration level is whether the paper is conceptual in nature or presents an implementation. Following this, Figure 1.7 shows the difference in level of integration for each application domain. For all application areas the level of integration is higher for conceptual papers than for case-studies. In aggregate 76.2% of conceptual papers propose true DTs, while only 42.9% of case-studies do. The difference is even more pronounced when counting the increase in Digital Models when moving from concept towards implementation; 16.7% of conceptual papers propose Digital Models while 46.9% of implementations are Digital Models. Please note that not all Digital Models are misidentified DTs as some authors choose to create a Digital Model first, before addressing the technical challenges to elevate it to become a DT. On Figure 1.7 a more granular overview is given. Notably the level of integration is decreasing for all application areas when going towards implementation. Paper from industrial applications achieve an overall high level of integration. The concept of DTs themselves stems from industrial research, and the community has an increased familiarity with the

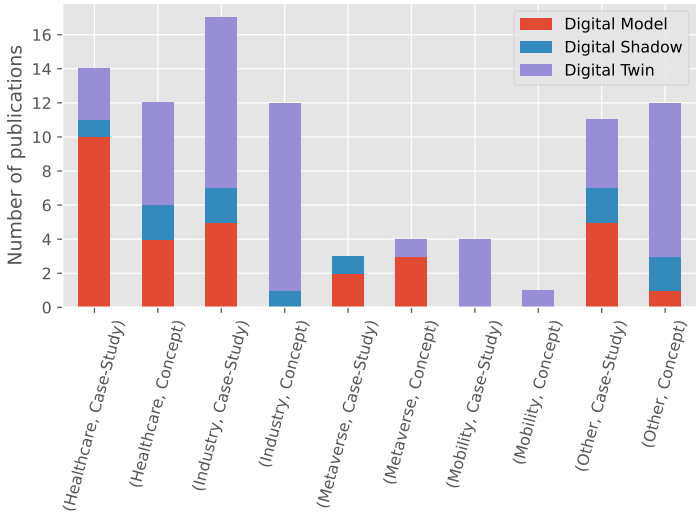


Figure 1.7: Level of data integration in reviewed paper, categorised by application domain and whether the study is conceptual in nature or a practical case-study.

concept. All the conceptual proposals are at least Digital Shadows indicating a stricter adherence to the tenants of digital twinning in the community. Nonetheless when implementing case studies nearly a third of the papers present digital models. Only Mobility applications succeed in having full data integration for both their conceptual models and implemented case studies, however the field is rather new and only 5 publications have been analysed. A third of papers from Healthcare that focus on the conceptual aspects of HDTs propose Digital Models instead, indicating a drift in conceptual understanding. Actual case studies are dominated by Digital Models because the research is mostly focused on creating working models first, and integrating said models later. Metaverse applications are highly focused on 3D avatar creation which tend to be decoupled from any human subject, consequently they should be considered Digital Models.

We observe that the interpretations and implementations of the DT tenants vary across application domains. The use of the word is ambiguous and often used interchangeably with Digital Model and Digital Shadow.

Twinned Entity

HDT is a special application of a DT. While a DT can in principle have any system as its twinned entity, an HDT should have a human entity as its twin. Understanding what is exactly meant by human entity constitutes the second part of our HDT definition.

From the literature a large majority of cases, to be precise 94.4%, considers a human individual as the twinned entity. However, there are two other notable categories that emerged; human roles, and systems containing humans. A breakdown of the exact count is shown in Figure 1.8.

For human roles a model is created that is meant to reproduce the behaviour or attributes of a human class or a human in a specific role. These models are not connected in real time to an individual nor do they seek to simulate a specific individual. All three examples have a low level of data integration, and all qualify as Digital Models. As an example, in Balaji et al. [2023a], a model is created that reproduces the behaviour of a control room operator based on a

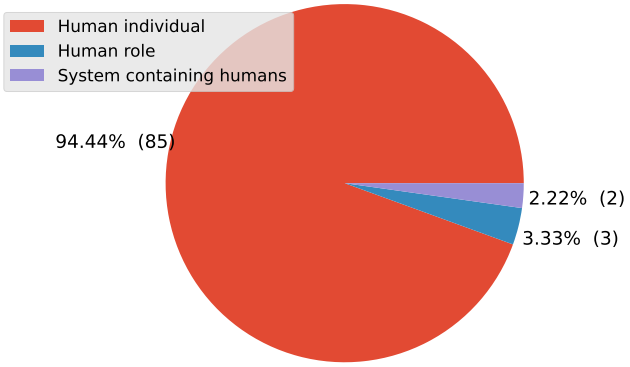


Figure 1.8: Count of publications categorised by the nature of the twinned entity.

cognitive architecture (ACT-R). However, there is no live data from an individual being fed into the model, the goal instead is to simulate and approximate real operator behaviour. In other work a 3D model of a human is created either while sitting [Prokscha et al., 2022b] or walking [Prokscha et al., 2022a] to determine the impact on wireless connections using raytracing. While data from humans is used to adapt the model, they remain generic in nature and don't seek to describe one specific individual.

The second category is DTs of systems containing humans. If the authors create a DT of a system that contains humans (e.g.: factory, household, traffic system, hospitals) and identify the whole as an HDT, then we count it in this category. In Ascone and Vanderhaegen [2022] the authors extend the concept of DTs with a special focus on humans. They have a particular interest in Human-Machine systems and propose a holistic DT for the entire system. In that context the human is entirely integrated as part of the larger DT, named *Human-Machine DT*, and does not have its own HDT. Interestingly two types of humans are distinguished; those who are subject to data collection and are modelled; and those that are users of the DT through the service layer. In Montini et al. [2022] an HDT reference model is proposed which includes factory entities and events as well.

From this analysis we can observe the overwhelming consensus that the twinned entity of an HDT is a human individual. It is to be distinguished from a human user of the HDT, which is similar to the operator of a traditional DT.

1.4 Challenges and Research Directions

1.4.1 Communication and Connectivity

HDTs rely on the data they receive but reliable data transfer is a major challenge. Depending on the use case, the twinned human individual may be moving in various environments throughout the day. Minimising the effect of mobility is a key technical challenge [Okegbile et al., 2023b]. Lacking network coverage can lead to gaps in data collection which directly impact the HDT's ability to run models in real-time and provide feedback to the twinned individual. A promising approach for continuous real-time synchronisation is prediction-enhanced connectivity which uses

the HDT's prediction to relieve the need for constant synchronisation [Okegbile et al., 2023a]. The computing resources to run the deployed models should be in physical proximity of the human to reduce latency. Techniques like cloud-edge computing are promising approaches [Wang et al., 2022b] that can help minimise the latency. Interesting choices have to be made for what models should be run at on an edge node and which should be delegated to the cloud. The place of models hosting, resource allocation and task off-loading are decisions that need to be optimized. A two-timescale optimization approach, where frequent decisions are run on the edge, while rare decisions are run on the cloud, can help balance accuracy, delay and energy consumption [Yang et al., 2024]. A set of guidelines to help designers decide which computation delegate to what level would be a helpful contribution to the community.

1.4.2 Sensing Human Data

Humans are complex systems and our capability to measure and model the human state is limited. External features like movement, position or gaze are measurable with current day technology through computer vision methods and eye tracking methods. Even for external features data from various sensor types are often required which leads to a need for effective and efficient data fusion techniques. Data about human physiology can be gathered through wearables but only provides a limited picture of the actual human state. Heavy limitations arise when trying to capture data about human cognition [Lauer-Schmaltz et al., 2022] where the measurements often rely on invasive imaging techniques like electroencephalography or functional Magnetic Resonance Imaging. Even wearables have been seen recognised as uncomfortable and intrusive if worn for long periods of time [Feng et al., 2025]. Last, not all human aspects can be gathered through sensors, making data directly provided by the twinned human (e.g. issued from questionnaires) important for an HDT. Depending on the use-case, complementing real-time observations from sensors with "manual" data becomes essential. For example, personality today can mostly be obtained from standardised psychological tests (e.g. Big Five [Goldberg, 1993] or the like). Attempts exist to deduce personality traits from behaviour in games [Naudet et al., 2015], but results are not as reliable as with normalised questionnaires. Another example is biographic data, which most of the time cannot be invented, though they can be retrieved from digital records when available.

Some important parameters, like cognitive load, are not directly observable but need to be inferred from other measurements. The current use of subjective questionnaires that are not real-time of interrupt work is a major limitation [Liu et al., 2025]. Methods to objectively measure cognitive load by integrating physiological signal analysis and subjective evaluation tools present a promising research direction to enrich the available information for the HDT[De Giglio et al., 2026].

1.4.3 Modelling

Once data is gathered about a specific aspect of the human individual, the challenge becomes modelling their state, evolution and behaviour. Current HDT developments are highly specialised and typically seek to create models that focus on very few aspects of the individual and are restricted to specific contexts. As a result, a large variety of models are built and often function as proof of concept. The ambitious vision of HDTs is to create one persistent DT that accompanies the twinned individual throughout their life and is able to provide support in a large variety of contexts. To work towards this vision the efforts across communities need to be coordinated such that a multitude of specialised HDTs can be integrated into a holistic one. A standardised

framework and architecture with precise design guidelines with the idea of integration in mind would be of enormous benefit to developers. Multiple frameworks and architectures have been proposed, often with a specific application in mind, but the current challenge is to build consensus across disciplines and decide on a standard.

Independent of the long-term vision, challenges with specialised models persist and need to be addressed. When creating a specific model, the human nature of the subject creates a set of challenges. As not all types of data can be gathered at all times, models need to be robust to data gaps. Similar robustness is needed for poor data quality, especially when complementary manual data entries are used in the model. Especially Machine Learning models, and in general data-driven models, are affected by this limitation as they need large volumes of high-quality data to be trained. The data scarcity and gaps may be addressed by Artificial Intelligence-Generate Content (AIGC). This approach advocates for the use of generative AI (e.g. GAN [AlAmir and AlGhamdi, 2022] or Diffusion model [Kazerouni et al., 2023]) to create synthetic data. First models are pre-trained on general datasets, before fine-tuning it to the specific use, like a rare disease or personalisation to a specific individual [Chen et al., 2024b]. In all cases trained data-driven models need to be validated in field conditions to determine whether they generalise well, especially when datasets are created in lab conditions. Explainability is an important requirement for critical application cases where an operator or Healthcare professional needs to understand the reason behind a given output. In more general terms explainability of the models is a key factor in building trust and acceptance.

Models may require multi-modal data to describe, track and predict the behaviour of an individual. In this context data fusion is a challenge for development [Pesapane et al., 2022], which directly impacts the models that can be built on top of fused data. The phenomena influencing the state and behaviour of a human happen at different time scales which may pose a challenge to modelling efforts. A slow process like ageing, or shifting moods, impact the behaviour, just like fast processes like recalling instructions, or recognising a change in the environment.

HDTs that model the interaction between the twinned human and environment require data about the surroundings to build good situational awareness. An example is fall simulations where the presence of obstacles influences real world behaviour which needs to be reflected in simulation model [Ferrari et al., 2025].

If the HDT requires real-time operations (e.g. in collaborative motion planning [Joseph et al., 2021]), or fast feedback, it imposes heavy restrictions on the computational complexity a model can have. Depending on the human aspects that are modelled the computational complexity for high-fidelity models can easily be prohibitive, therefore a trade-off between fidelity and complexity is an inevitable dilemma that arises. Cumulative time delays from model execution and data transmission may cause negative effects in HRC setups [Wang et al., 2024c; Feng et al., 2025], especially for computationally costly deep learning models.

Cognitive aspects represent a major modelling challenge [Lauer-Schmaltz et al., 2022]. HDTs seek to emulate humans including their cognition and could take inspiration from Cognitive Digital Twin (CDT) research. CDT development aims to create a digital representation of a physical system with additional cognitive capabilities to perform autonomous activities and adapt in real time to varying operational contexts [Zheng et al., 2022]. The field has advanced modelling of perception, memory and reasoning. In particular CDTs can be driven by cognitive architectures that recreate the cognitive cycle of human cognition, to enable real-time adaptation of technical systems [Ali et al., 2024; Al Haj Ali et al., 2024b]. The CDT cognitive cycle may well be used to represent human cognitive processes dynamically on different time scales and in different interaction contexts [Ji et al., 2023]. However, CDT and HDT pursue different goals,

since the former seeks to use cognition to enhance a DT functions (e.g. analysis), while the latter needs to emulate cognition of a human-being [Naudet et al., 2024].

1.4.4 Societal and Ethical Challenges

Technology Acceptance Model (TAM)

With new technologies come questions about their ethical use and societal acceptance. HDTs are inherently built on personal data that require more stringent protection. Thus, HDTs must comply with regulatory frameworks such as the European General Data Protection Regulation (GDPR) [Fratta et al., 2026]. Poor data security & privacy, faulty models and lack of explainability can all erode trust in HDTs and hamper societal acceptance.

Currently, acceptance of HDTs is not evaluated systematically as the research field is young and the maturity of implementations remains limited. However, as HDTs mature their acceptance should be analysed following appropriate methodologies. From general technology acceptance model such as the Technology Acceptance Model (TAM) [Davis et al., 1989] or Unified Theory of Acceptance and Use of Technology (UTAUT) [Venkatesh et al., 2003] provide valuable guidance to evaluate the acceptance of an emerging technology. A major drive of acceptance is the expected and perceived performance resulting from using the technology. Given the special relation between the twinned human and HDT, it may prove valuable to develop a specialised acceptance model for HDTs or revise existing models for this special case.

For secure data sharing, blockchain technologies have been identified as a solution but struggle to meet latency requirements because of their expensive consensus mechanisms. A potential solution is dividing the blockchain into shards for parallel processing and use a reputation-based mechanism to append transactions to the blockchain [Okegbile et al., 2024]. When building data-driven models classical approaches centralise the data for model training which necessarily requires the transfer of sensitive data. To maintain privacy federated learning is popular solution, i.e., instead of sharing data directly, a local model is trained and shared to be subsequently aggregated into a general model. The shared local models may still be used to infer personal data. Adding calibrated noise to the local model parameters has been proposed as a solution [Okegbile et al., 2023c]. Additionally, Okegbile et al. [2023c] propose a proof of model quality mechanism to validate the local models before aggregation.

Beyond challenges about trust that are related to technical challenges, there are ethical questions that need answering. Concerning data collection HDTs are potentially unduly intruding into an individual's private life, which alongside the risk of data breaches can lead to catastrophic consequences. A potential problem is informing the user about the functioning of the HDT and the use of their personal data to get informed consent for the data collection and processing. A clear process is needed for revoking consent and deleting the data in question. Linked to data deletion is the question of what should happen with an HDT upon death of the twinned individual. The gathered data can be of great benefit to domains like medical research, but it is unclear whether the data of the deceased can be used ethically after their death [Iqbal et al., 2022]. When HDTs rely on data-driven models, the problem of bias and discrimination becomes significant [Hernandez-Boussard et al., 2021], especially when the HDT outputs are used for critical decisions such as who to hire or what treatment to apply for a disease. It can be argued that the twinned individual should retain control over any decision that impacts their own life. If HDTs become a tool to empower its users equitable access becomes an important priority to avoid creating a two-tiered society with a significant section of the population left behind [Lonsdale et al., 2022]. A last problem to consider is the intentional misuse of HDTs which includes surveillance by governments or corporations, manipulation by political operatives

or commercial marketing.

HDTs concentrate data about a human individual and with an appropriate design philosophy can give the individual increased control over their data in question and how it is used, including what decisions impacting their own life can be made with it. Establishing guiding principles for ethical HDT development are an urgent need for industrialists [Cardin and Trentesaux, 2022] and requires interdisciplinary work with researchers from social sciences.

1.5 Discussion

In our systematic literature review, we surveyed the current day understanding of Human Digital Twin (HDT) across all application domains. We have shown how the concept started in Industry before spreading to Healthcare and subsequently to an ever increasing number of new domains. With new domains came new viewpoints, which contributed to a divergence of conceptual understandings. From the analysed papers, we extracted how HDTs were conceptualised and observed a consensus that an HDT's twinned entity is a human individual. In contrast, researchers have different views on the data flows between the individual and their HDT.

HDTs have the potential to provide valuable, personalised support to its users and enable a new generation of human-centric systems for Industry 5.0. However, HDT development faces a multitude of technical challenges and ethical implications. Technical challenges revolve around gathering data from humans, transferring them and modelling the twinned human individual, especially their behaviour. Measuring internal states like cognition remains an enormous hurdle. Current developments mostly take the shape of proof-of-concepts that operate in a highly specific context, but the long-term vision is creating a persistent Human Digital Twin to accompany the user in various life scenarios. Beyond technical challenges HDTs have profound ethical and societal implications. Trust and societal acceptance depend on stringent data security, highly reliable models, and explainability of their outputs. The users need to provide informed consent before their data is used, beyond which data privacy needs to be guaranteed.

1.6 Contribution Positioning

Within the context of Human Digital Twin research and their application, this thesis seeks to contribute to the following specific challenges and contemporary limitations:

Limitation 1: *Ambiguity about the essential elements of Human Digital Twins.*

The literature review showed how the concept of HDT has spread to an increasing number of new domains. However, the concept itself is not understood uniformly across, or even within, fields. While a general consensus on the twinned entity exists, the interpretation of what defines a Digital Twin varies. The words *Digital Twin*, *Digital Shadow* and *Digital Model* are used interchangeably leading to ambiguity. This leads to the misidentification of HDTs and disregard of their key fundamental properties and requirements. Therefore, a need for a thorough disambiguation exists such that HDTs can be clearly defined and categorised.

Limitation 2: *Lack of a unified meta-model for Human Digital Twin development.*

HDT research is rather young leading to implementations and model developments that are focused on narrow use cases. Due to the technical challenges, but also the conceptual ambi-

guities, most HDT developments stop short and present *Digital Shadows* or *Digital Models* as results. While a few general frameworks, architectures and meta-models have been proposed, there remain points of disagreement. Notably, whether HDTs are purely digital in nature or should be embodied is not clear. The lack of a recognized general model harms re-usability of developed architectures and models.

Limitation 3: *Approaches to emulate cognitive aspects are lacking.*

Cognitive aspects have been identified as major obstacle to detect and model. In contrast to machines, the direct detection of a human's internal states that drive their behaviour is uniquely challenging. Consequently, the development of HDTs that account for the twinned human's behaviour is hampered. Alternative methods, in particular cognitive architectures, simulate observable behaviour based on modelled internal states. However, the possibility to use these models for real time emulation of an individual's behaviour remains unexplored.

In line with the identified limitations, this thesis sets the following objectives:

- Investigate and analyse the fundamental properties of HDTs to propose a general domain-agnostic definition and categorisation system.
- Develop a meta-model for HDTs defining its internal components and interfaces. Additionally, the model includes the relation to external systems such as the twinned human.
- Formalise the properties of emulation models for human cognition and define the requirements such a model needs to fulfil for effective emulation.
- Instantiate a cognitive architecture to function as an emulation model for a human performing an assembly task.

Chapter 2

Human Digital Twin Fundamentals

2.1 Introduction

This Chapter presents the core conceptual contribution of the thesis. From the literature review from Chapter 1, we identified limitations of Human Digital Twin research that should be addressed. The concept of Human Digital Twin is used across a large variety of domains leading to contradicting understandings of the concept. To clarify this ambiguity this thesis develops and proposes the following:

1. A general domain-agnostic definition and classification system for Human Digital Twin.
2. Development of a meta-model anchored in General Systems Theory (GST) [Von Bertalanffy, 1972]

With the definition and systemic meta-model, we seek to provide an answer to (**RQ1**): *What defines a Human Digital Twin and what are its fundamental elements?* The systemic meta-model provides the conceptual framing for the work presented in later chapters.

This chapter draws from two publications made as part of this thesis.

- Gaffinet et al. [2025a] presents a systematic literature review to propose a general definition and classification system for HDTs.
- Gaffinet et al. [2025c] presents a systemic meta-model for HDTs.

2.2 Classification & Definition

Through our literature review in Chapter 1 we have explored the main conceptual components of HDTs as they are understood and used by various communities. It revealed that the term HDT is not only used to refer to true DTs, but also to Shadows or Models. A good consensus only exists for the nature of the twinned entity, which is a human individual. The seemingly interchangeable use of the key terms like Digital Twin, Digital Object or Digital Model requires clarification. In Kritzinger et al. [2018] the notion of Digital Object and Physical Object is introduced as well as a clear distinction between models, shadows and twins. To avoid confusion during our analysis section, Figure 2.1 summarizes the vocabulary used for DTs and HDTs and common synonyms for the various entities. In general, a *Digital Twin* has exactly one *Twinned*

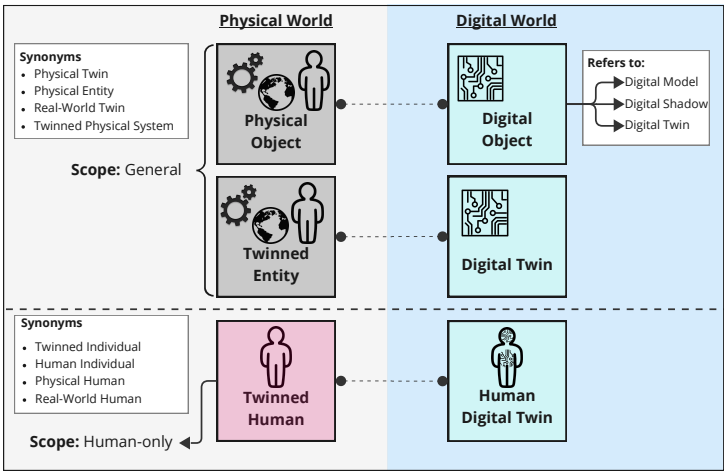


Figure 2.1: Vocabulary for Digital Twin and Human Digital Twin.

Entity. The twinned entity is also commonly referred to as *Physical Twin*, *Physical Entity*, *Real-World Twin* or *Twinned Physical System*, all of which are synonyms. Kritzinger et al. [2018] use *Physical Object* and *Digital Object* to refer to a wider class of entities. The *Digital Object* keyword is used to refer to *Digital Twins*, *Digital Shadows* or *Digital Models*. For *Human Digital Twins* the scope of the twinned entity is necessarily a human and is commonly referred to interchangeably as *Twinned Individual*, *Human individual*, *Physical Human* or *Real-World Human*.

When surveying the literature in Chapter 1 we based categorized HDTs presented in papers based on their level of data integration as in Kritzinger et al. [2018]. For a more detailed explanation on *Digital Models*, *Digital Shadows* and *Digital Twins* refer to section 1.3.3.

We advocate for the adoption of this established categorization for HDTs to rectify the conceptual ambiguity we identified during our systematic literature review. Since an HDT is a type of DT, it has to inherit the same fundamental requirements for automated data integration to be considered a "true" DT. The additional categories enable a clear distinction between simulation models (*Digital Models*), monitoring systems (*Digital Shadows*) and more complex bi-directional architectures (*Digital Twins*). We hope that this disambiguation ensures that important keywords carry clear meaning and minimizes the misidentification of models for HDTs.

In Figure 2.2 we give a basic overview of the differences between ***(i)*** Human Digital Model; ***(ii)*** Human Digital Shadow; ***(iii)*** Human Digital Twin and; ***(iv)*** traditional Digital Twin. A traditional DT receives data automatically from a technical system through a data acquisition interface. It processes the data and runs models to return automatic feedback to the twinned system. Likewise, an HDT receives data automatically from a human individual through a data acquisition interface. However additional data can be entered manually from the twinned human. Contrasting with Traditional DTs that may only receive manual entries from, for example operators, but not directly from the technical system. The data is processed and models are run to create feedback that is returned through a human-machine interface to the twinned human. How the feedback is given represents a major difference from traditional DTs. Feedback to a technical system includes direct commands that are processed and executed according to their intended design. However, humans cannot and should not be controlled like machines, thus the feedback takes the shape of guidance, suggestions and maybe even orders. But it does not come equal to a command that is executed deterministically. More details on feedback are discussed

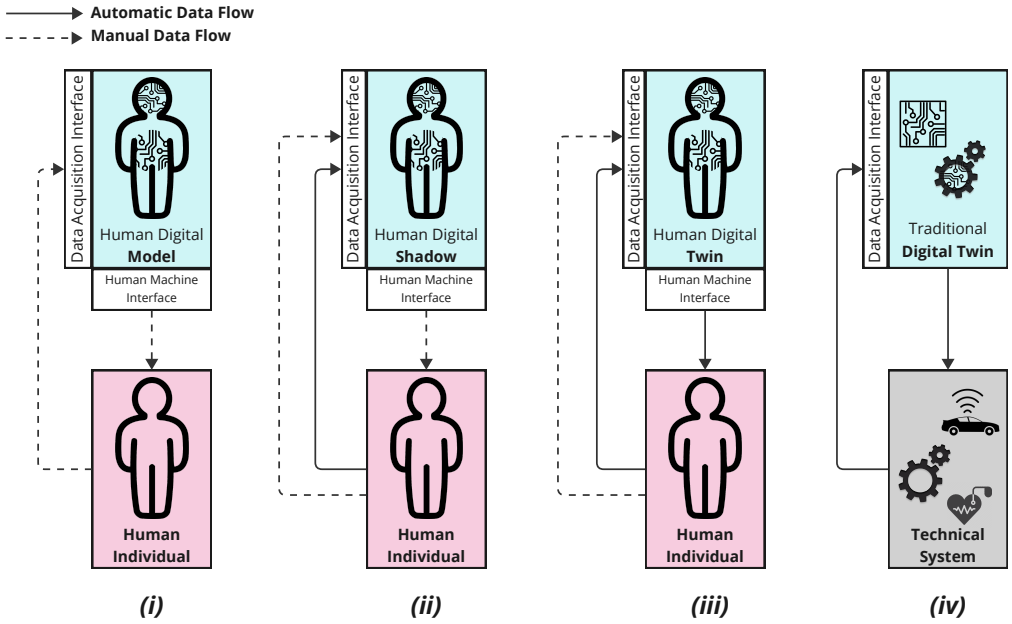


Figure 2.2: Overview of entities and data flows for ***(i)*** Human Digital Model; ***(ii)*** Human Digital Shadow; ***(iii)*** Human Digital Twin and; ***(iv)*** traditional Digital Twin.

in a dedicated Section 2.2.4.

Each of the elements is discussed in detail in the Sections below; ***(i)*** Physical Object, also referred to as twinned entity (Section 2.2.1); ***(ii)*** automatic data flow to the Digital Object (Section 2.2.2); ***(iii)*** Digital Object (Section 2.2.3); ***(iv)*** automatic feedback to the Physical Object (Section 2.2.4). After covering all the elements that make up an HDT, and laying out the differences with a traditional DT, we aggregate all our findings in a general and application agnostic definition (Section 2.2.5).

2.2.1 HDT Specificities: Physical Object

For DTs in general the twinned entity can be any physical object. One aspect that separates an HDT from a traditional DT is that the twinned entity needs to be human. In the preliminary definition the twinned entity has to be of *human nature*. However, in Section 1.3.3 of the literature review, we have established that there is high consensus that the twinned entity needs to be a *human individual*. However, two alternative minority viewpoints exist. The first, states that an HDT can represent a *class of humans* or *human role*, for instance the role of factory worker. This view implies that an HDT does not have a unique relationship with its twin. Instead, it can be connected to different humans to provide generic feedback. The second minority viewpoint is that an HDT can have a system that contains human elements as twinned entity. An example is a shop floor with workers that is represented in a single DT. Both cases break the unique one-to-one relationship between the HDT and its twin.

We align ourselves with the consensus and refine the preliminary definition with the necessary, but not sufficient, condition that the twinned entity is a human individual.

2.2.2 HDT Specificities: Data Flow to the Digital Object

When considering traditional DTs, for a machine, or an autonomous car for example, there is the possibility to design the technical system with integrated sensors. For a human individual does not come with integrated sensors and some of the internal processes are not easily measurable with today's technology. Lauer-Schmaltz et al. [2022] identify that in particular gathering information about human cognition represents a substantial research gap. Besides the challenge of gathering data that provides insights into the human's internal state, it is important to consider how the data is transmitted to the HDT i.e., whether the data flow is manual or automatic, and whether the data is received in real-time.

For traditional DTs and Digital Shadows, the data flow from physical object to digital object needs to be automatic [Kritzinger et al., 2018]. The same necessary condition applies to HDTs and can, for example, be fulfilled with data collection through wearables or live video feeds. Depending on the application case the automatically gathered data may be insufficient and complementary manual input may be required. An illustrative case study is the prediction of fitness outcomes based on manual inputs about food consumption and automatically received data from wearables [Barricelli et al., 2020]. It is worth noting that traditional DTs may also receive manual inputs, but from an operator, not the twinned entity. This is markedly different from receiving manual data directly from the twinned human. An analogous situation for HDTs would be manual inputs from a Healthcare professional that has the role of monitoring the twinned human, similar to an operator monitoring a twinned technical system.

A second point of contention is whether a DT, in general, needs to receive data in real time as explicitly stated in the definition of DT in Semeraro et al. [2021]. The ISO standard omits the real time requirement and defines DT as *"fit for purpose digital representation of an observable manufacturing element with synchronisation between the element and its digital representation"* [ISO-23247, 2020]. A real time requirement corresponds to a high minimum sampling rate. However, for both traditional DTs and HDTs, the minimum sampling rate is dependent on the application case. To remain general, the sampling rate should be fit for purpose, which includes both the cases that require real-time processing and those that have more relaxed time constraints

We align ourselves with the ISO standard's point of view that the HDT needs to be *fit for purpose* and thus the application date dictates the required update frequency. For a Human-Robot Collaboration case where collisions have to be avoided it is necessary to track and process the movement of the human in real time. But to monitor the development of a disease to adapt treatment a low update frequency may be sufficient to achieve effectiveness. Nonetheless, both cases deserve to be considered HDTs which is why we believe the *automatic* nature of the data transfer is the relevant component of the definition.

In short, we take the position that an HDT necessarily receives data automatically from its twinned human individual, but it is not necessary to receive the data in real time. Instead, the sampling rate needs to be fit-for-purpose. The automatically received data can, but does not have to, be complemented by manual entries.

2.2.3 HDT Specificities: Digital Object

The Digital Object refers to the HDT itself and is purely digital in nature. Its primary role is to be a digital representation of the twinned individual, requiring a model that emulates the state, and possibly behaviour, of the twinned human. Like for a traditional DT, this model may be used to predict future states or run simulations. To build and run its model, the HDT receives

data from its twinned individual through deployed sensors and manual inputs. The personal nature of the data makes security and data management a fundamental responsibility of the HDT. In addition, an HDT may deploy communication interfaces to receive data from other DTs or technical systems to provide the needed context to the hosted models and send relevant messages to surrounding systems. The key components described here are similar to traditional DTs, nonetheless we see two challenges for HDTs, namely the complexity of humans and the highly dynamic context they live in.

Traditional DTs mostly have human-made technical systems as their twinned entity. The engineering efforts to develop the technical system establish a minimum level of understanding and mastery of the system at hand. While technical systems can be highly complex, they pale when compared with the complexity of a human individual which is composed of highly interlinked cognitive, physical, physiological, social and biographic aspects. To truly reflect the state of a human, let alone predict their evolution, each aspect would need to be modelled including their interactions and dependencies. For practical use cases it may be sufficient to model a subset of human aspects to achieve a fit for purpose digital representation. A second challenge is the context in which a human or a technical system lives or operates in. Most technical systems are designed to operate in highly specific contexts, and their DTs can be effective by hosting highly specialised models e.g.: autonomous cars operate on roads, and industrial robots are operated on assembly lines. In contrast, a human individual is exposed to an enormous spectrum of contexts within their lifetime making the transferability and generalisation of models increasingly challenging.

HDT development can be framed along two distinct levels. The first level involves creating context-specific models designed to provide support within a limited scope. An example is a muscle fatigue HDT that is temporarily connected to its twin to provide assistance during a specific physical task [You et al., 2025]. While these task-specific HDTs establish a unique and tightly integrated relationship for the duration of the interaction, they are not persistent, and their underlying models are often not generalised to other scenarios. This contrasts the second level of framing: the long-term vision of the HDT. In this paradigm, an HDT remains persistently connected to its twinned individual throughout their life, hosting a suite of generalised models to provide holistic, cross-contextual support. The primary challenge in this level is the creation of a set of interconnected models that generalize well across diverse environments and can leverage each other's outputs as predictive inputs. For instance, a generalized physiological model could continuously track states such as Heart Rate Variability (HRV), heart rate, and Galvanic Skin Response (GSR). These outputs can then serve as inputs for diverse modules; for example, a cognitive load model could exploit GSR data to infer mental strain, while a muscle fatigue module could utilize heart rate metrics to refine its predictions of physical exhaustion.

Within the AI4C2PS project, we developed and published an ontology for human cognition and abilities [Naudet et al., 2025; Zehnder and Naudet, 2025]. The resulting HUman ONtology (HUMO) draws from perspectives from neuropsychology, cognitive informatics and cognitive architectures to formalise a high-level model for any human dimension further detailing cognition and ability-task relationships. This ontology-based approach provides a common vocabulary for integrating generalised models about any human aspect.

In summary, an HDT primarily emulates its twinned human's state and potentially behaviour. Current HDT developments are constrained to task specific models and support functions. However, in the long-term, HDTs may be continuously enriched with generalised and interconnected models to provide holistic support throughout the twinned individuals life. Beyond emulation the HDT may have additional functions such as simulation or prediction. Personal data needs to

be safeguarded, and contextual information may be received by other DTs or technical systems.

2.2.4 HDT Specificities: Feedback to the Physical Object

According to Kritzinger et al. [2018], *"A change in state of the physical object directly leads to a change in state of the digital object and vice versa."*. This means that for classical DTs, feedback should be direct and automated. However, feedback is partially directed towards the twinned system (*prescribing real time actions* [Semeraro et al., 2021]) and partially towards the human operator to adapt their decisions when optimising operations or scheduling maintenance. So we can distinguish two kinds of feedback: *(i) direct feedback* addressed at the twinned system directly, and *(ii) indirect feedback* which goes through an operator before any changes to the physical object are applied. The latter kind inhibits the automatic application of the feedback.

For traditional DTs the form of the feedback varies from warnings to the operator to direct control of the twinned entity. An illustrative example is from predictive maintenance where the DT pro-actively schedules maintenance based on real-time data of twinned machinery and the estimated Remaining Useful Life [Aivaliotis et al., 2019]. A DT may simulate a robotic arm to optimize path selection and provide a program with optimized behaviour as feedback. Even direct control is possible with DTs, for example the ventilation and air conditioning of buildings may be directly controlled to maintain ideal working conditions. In contrast a human cannot be directly controlled by their HDT. Instead, recommendations and warnings can be sent, but a human may miss or ignore feedback messages. For instance, a recommendation to take a break during an industrial assembly task can easily be ignored by the worker.

Just like for technical systems, humans can receive *direct* and/or *indirect* feedback from their HDT. Direct feedback is necessarily mediated through a Human-Machine Interface while indirect feedback goes through a third system before impacting the twinned individual. To fulfil the condition to be a DT the feedback needs to be automatic, which is the lens through which we analyse both direct and indirect feedback below.

1. ***Direct feedback*** is necessarily mediated through a Human-Machine Interface to be understandable to a human individual. Any model output from the HDT needs to be translated into a human understandable form and can take various forms, e.g., visual feedback on a screen, push notifications, lights and sounds for alerts or even haptic feedback. Depending on the modality of the feedback the human can easily ignore or unwillingly miss it, and even if the feedback is successfully received it can be disregarded by the human. Overall, the feedback can be considered weak in its potential to control the twinned individual which is a fundamental difference to traditional DTs. In our view, any direct feedback that is automatically sent, independently on whether it has been accepted, fulfils the condition of Kritzinger et al. [2018].

2. ***Indirect feedback*** goes to another system first which can be a human entity or a technical system. An illustrative example for indirect feedback through a human is a Healthcare professional. They may use an HDT to receive updates about the progress of an affliction and adapt the treatment based on it, much alike the case of an operator receiving information about a machine to adapt its operating parameters. The introduction of a human entity that has to manually apply any change to the twinned human's state breaks the automatic nature of the data transfer and should not be considered an HDT. A technical system can be weakly or strongly coupled. A strong coupling implies that a change in the technical system will lead to a change in the state of the twinned human as well. Thus, feedback

received by a strongly coupled systems is functionally the same as direct feedback. In contrast, feedback mediated through a weakly coupled system is indirect in nature.

An HDT needs to provide automated feedback to the twinned individual. Direct feedback through a Human Machine Interface, initiated automatically by the HDT, fulfils this condition. Indirect feedback through another human entity, or weakly coupled technical system, is insufficient and would fall into the category of Human Digital Shadow.

2.2.5 General Human Digital Twin Definition

Based on the prior sections on the specificity of an HDT compared to a traditional DT we synthesise a general domain-agnostic definition. From the prior discussion the following elements have been identified:

1. The twinned entity is necessarily a human individual.
2. The input data is necessarily gathered automatically and may be complemented by manual data entries.
3. The HDT necessarily needs to host models to reflect the state of the twinned individual, that is be able to emulate the twinned individual. The HDT may host additional functions such as simulation or prediction.
4. Feedback is necessarily given automatically to the twinned human.

Proposed General Definition: Human Digital Twin

A Human Digital Twin is a class of Digital Twin whose twinned entity is a human individual. It is a digital representation of the twinned human, emulating their state and dynamics. It automatically receives data, with optional manual entries, and provides automatic feedback directly to the twinned human.

With this definition we give our best proposal that respects the tenants of digital twinning and duly considers the conceptual understanding from different disciplines. The definition clarifies the exact nature of the twinned entity and characterises the required data transfers. It also requires HDTs to emulate their twinned human. HDTs may have additional functions, such as simulation and prediction functions. Also, communication with other DTs or technical systems is allowed but not required.

2.2.6 Special Case: Augmented Human Digital Twin

There is a special case which tests the boundary of this condition. Is it still an HDT if the twinned entity is a human individual including a strongly coupled technical system? A system is strongly coupled when its components are highly interdependent and hard to separate. A hypothetical example would be a pacemaker or insulin pump where the twinned human is highly dependent on the technical system's functioning. In contrast, for weakly coupled systems the human can remove themselves from the interaction and exist independently. An illustrative example of a weakly coupled technical system is a smart car that provides driver assistance [Hu et al., 2022]. Exoskeletons are another great example, where the coupling can be weak if a worker is using

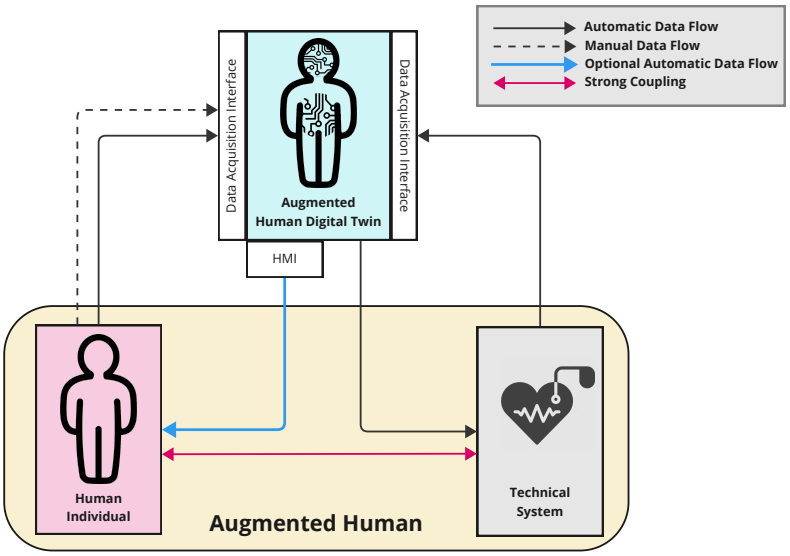


Figure 2.3: Data flows of an augmented Human Digital Twin (aHDT) where a human and a strongly coupled technical system are twinned to a single DT.

the exoskeleton temporarily [Constantinescu et al., 2019], or strong if the augmented human is permanently dependent on the system. We propose the term *augmented Human Digital Twin* (aHDT) to refer to a DT encompassing a human with a strongly coupled system. Figure 2.3 shows the dataflows between the involved entities. The properties are discussed following the same structure as our general HDT analysis:

- The **Physical Object** is a human with a strongly coupled technical system. Changes in the technical system directly impact the human and vice versa. Additionally, the link between human and technical system is persistent, meaning the human cannot easily choose to remove themselves from interacting with or depending on the technical system.

- **Data Flows to the Digital Object** come from both the human and/or the technical system. The technical system can communicate data about its own state as well as measurements about the human state. Additional data about the human can be provided by other sensors, for instance by wearables.

- The **Digital Object** gathers data from both entities to track and emulates the joint state and dynamics of the human and technical system. The aHDT's model can, among others, be used to run simulations and make predictions about future states.

- **Feedback to the Physical Object** can be sent to the human or technical system. As in a regular HDT feedback to the human takes the form of guidance and suggestions as no direct control is possible. However, feedback to the technical system can include commands for execution. Due to the strong coupling any commands executed by the technical system directly impact the human.

Proposed Definition: Augmented Human Digital Twin

An Augmented Human Digital Twin is a class of Human Digital Twin whose twinned entity is a human individual coupled strongly with a technical system. It returns automatic feedback directly to the twinned human individual or to the technical system.

2.2.7 Summary of Human Digital Twin Fundamentals

To conclude this Section, the proposed conceptualization of HDT addresses the ambiguity identified in the literature by establishing a rigorous foundation based on two primary dimensions; ***(i)*** the *nature of the twinned entity* and; ***(ii)*** the *level of data integration*. Figure 2.4 synthesizes our findings. An HDT is formally defined as a class of Digital Twin whose twinned entity is a human individual. It functions as a digital representation that emulates the individual's state and dynamics, characterized by *automatic data acquisition* (which may be supplemented by manual entries) and *automatic feedback* provided directly to the human. To further clarify the landscape, this framework allows for the naming of related concepts based on their specific properties:

- **Human Digital Model:** Characterized by manual data input and manual feedback, typically used for offline simulations.
- **Human Digital Shadow:** Features automated data flow from the human to the digital object but lacks automated feedback.
- **Augmented Human Digital Twin (aHDT):** A specialized class where the twinned entity is a human individual strongly coupled with a technical system, such as a pacemaker or an exoskeleton, receiving feedback to either the human or the technical system.

The proposed domain-agnostic definition and classification system is the first contribution ***C1*** of this thesis and provides the first part of the answer to **RQ1:** *What defines a Human Digital Twin and what are its fundamental elements?* This conceptual clarification is the foundation for the development of a systemic meta-model in the next section.

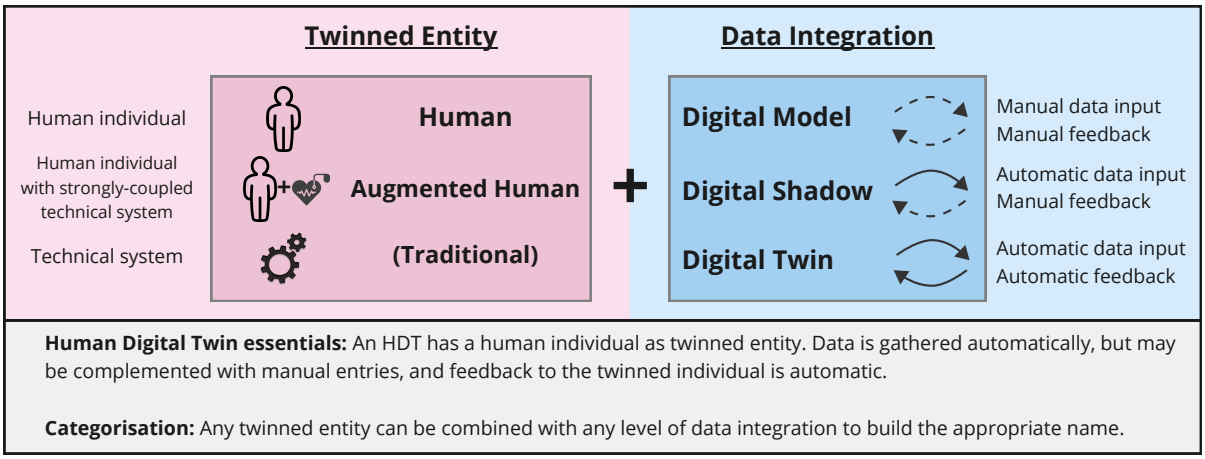


Figure 2.4: Essentials of Human Digital Twins and categorisation framework based on the twinned entity and level of data integration

2.3 Systemic Meta-Model

With our proposed definition, the fundamental aspects of HDTs are clarified. However, to guide future developments and facilitate reusability, a unified framework is required. The relationship between the physical and digital world is at the core of DTs. That includes the data sources about the physical world, and feedback mechanisms by the DT. For HDTs various models seek to clarify the relationship between these elements, but there is no consensus on the place of data sources and feedback mechanisms. He et al. [2024] propose a framework placing the data sources within the Physical Twin and feedback interfaces within the Virtual Twin. In contrast, Lin et al. [2024] consider feedback interfaces as separate from the HDT but assign data sources to the human. Lauer-Schmaltz et al. [2024] explicitly integrate feedback interfaces into the HDT and present a vision of embodied HDTs. Miller and Spatz [2022b] introduce a single interchange component for all data flows between the physical and digital world, providing no clear distinction. Layer-based architectures tend to not specify whether data collection and feedback mechanisms belong to the HDT [Wang et al., 2024b; Shengli, 2021b]. Beyond the need for conceptual clarification, most HDT models focus on isolated representations of individuals, disregarding their integration into a larger digital ecosystem. In contrast, high-level layer-based architectures include the larger ecosystem but provide little guidance on how to separate the various physical and digital entities to manage their interactions. We seek to address these short comings by developing a meta-model anchored in GST. Throughout this chapter we use the UML notation defined in ISO/IEC 19505-2:2012.

2.3.1 Core Entities

As the meta-model we develop should be coherent with the given definition, a first necessary condition for an HDT is that its twinned entity is a human individual. An HDT has a unique relation with their twinned human, specifically models, and support this individual. The objective of the HDT is very similar to that of traditional DTs, as both seek to emulate their twinned entity. Although creating traditional DTs is in no way trivial, the complexity of a human exceeds that of technical systems. Modelling the numerous physiological, cognitive, and mechanical processes and their dependencies is a major roadblock for holistic HDT development.

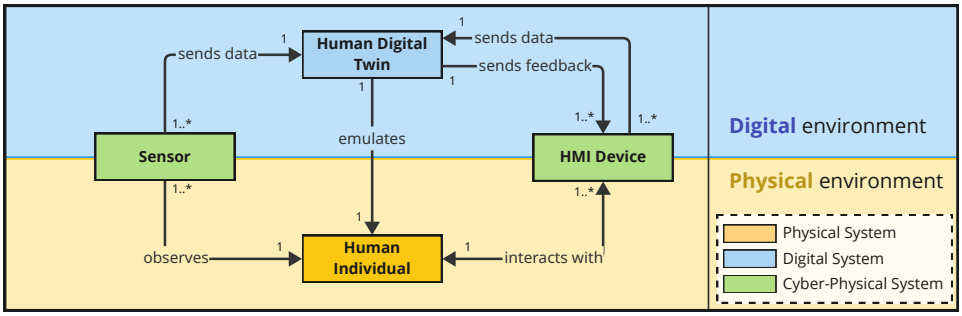


Figure 2.5: The *Human Digital Twin* - *Human Individual* pair, and the essential entities to enable emulation of the human by the HDT and interaction between both.

The previously mentioned personal nature of the data makes data security and privacy a duty of the HDT. Explainable and reliable models are an important requirement to avoid negative outcomes for the twinned human and build trust in the HDT.

Like any DT, an HDT needs suitable interfaces to implement the bi-directional dataflow with the twinned human. Those interfaces could be part of a DT; however, if we consider it as a pure digital system, they should not. In the case of the HDT, we have two additional entities involved in the DT-Physical Twin communication flow: *Sensors* and *Human-Machine Interaction (HMI) Devices*, that both serve as a bridge between the physical and digital world. *Sensors* gather data about the twinned human automatically and send them to the HDT. *HMI Devices* enable interactions between the human and their HDT. Feedback from the HDT, as well as manual input from the human, is mediated through this component. It may take the form of a tablet or screen and can loosely be referred to as the human-machine interface.

From the general definition emerge four essential entities to create an HDT, namely ***(1)*** *Human Individual*; ***(2)*** *Sensors*; ***(3)*** *HMI Devices*; and ***(4)*** the *Human Digital Twin* itself that is deeply dependent on the three other entities. Figure 2.5 illustrates these elements including their relations. The *Human Digital Twin* emulates the twinned *Human Individual*. One or multiple *Sensors* observe the *Human Individual*, produce data about them and send it to the *Human Digital Twin* enabling the required automatic data collection. The *HMI Devices* enable automatic feedback and additional manual input that the human may wish to provide.

Figure 2.5 illustrates the digital and physical environment using a coloured area. The four main entities are colour coded and categorised as *Digital*, *Physical* and *Cyber-Physical Systems* respectively. In the next Section, we introduce the systemic approach and the rationale for these categories.

2.3.2 Systemic Grounding

Von Bertalanffy [1972] introduced General Systems Theory (GST) which originated in the field of biology to address the complexities of biological organisms. Despite its origin, the theory defines core principles and concepts that are applicable to any domain. Since then, the theory has found success in numerous domains, including engineering and ecology. As a general theory, it goes back to the fundamentals of the complex objects that make our environment, their structure, and their dynamics, including not only biological systems like humans but also engineered systems like cyber-physical systems and digital twins. For this reason, it provides a strong ground for building a generic meta-model for digital twins, and particularly HDT.

Figure 2.6 illustrates the components we retain as a basis for a system in its environment

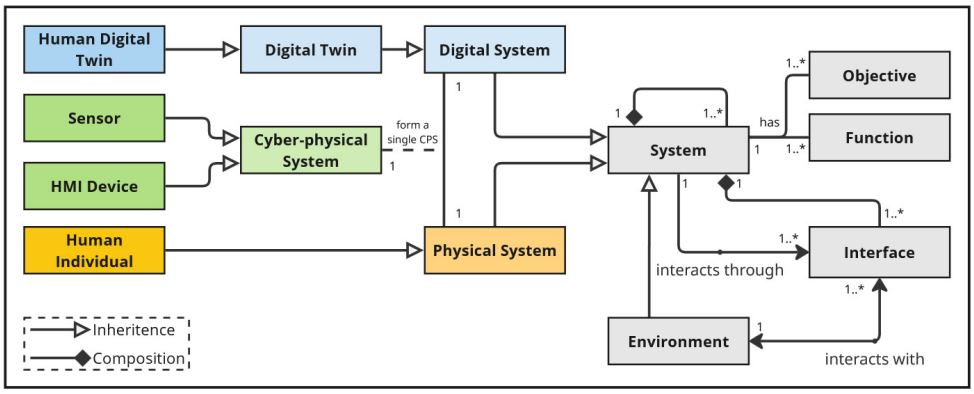


Figure 2.6: The four essential elements for HDT development as different types of systems, and the main elements of a systemic meta-model.

and a categorisation of the systems involved in HDT development. A *System* is defined by its components which interact with each other forming a coherent whole in an environment in which it acts. These components may be systems themselves. At its boundaries, it interacts with its *Environment* through *Interfaces*. The environment is itself a system, like the interfaces. Additionally, a *System* has *Objectives* and *Functions* to realise them. [Le Moigne, 1990]. Throughout the paper, we use the same three colours as a shorthand for *Digital Systems* (blue), *Physical Systems* (yellow) and *Cyber-Physical Systems* (green), while all other elements are shown in a default grey colour.

Each of the HDT's essential elements can be conceptualised as a system, with its own internal dynamics, interfaces, functions, and objectives. An HDT is defined as a digital representation and exists exclusively in the digital world. For conceptual clarity, we define *Digital System* as a *System* that consists exclusively of digital components. Both traditional DTs and HDTs are *Digital Systems* following this definition. Analogously, a *Physical System* consists exclusively of physical components, making the twinned human individual a purely *Physical System*. The two remaining systems, *Sensors* and *HMI Devices*, have the role of transforming physical signals into digital signals and vice versa. Bridging the digital and physical world can be considered as their objective. To achieve this, they must be able to simultaneously interface with *Physical Systems* (*Human Individuals*) and *Digital Systems* (*Human Digital Twins*). Hence, they necessarily have physical and digital components at the same time. This type of system is referred to as *Cyber-Physical System* (CPS) (see Lezoche and Panetto [2020] for a meta-model). A CPS is composed of a physical component and a cyber component, characterised by a high level of integration. In the context of this paper, cyber components and digital components are treated as synonymous.

2.3.3 Human Digital Twin Meta-Model

This Section presents the HDT as a *Digital System* by describing its digital components, interfaces, objective, and functions. The objective of an HDT is to emulate the twinned human and provide feedback to support them. Three digital modules are needed to achieve this objective. A *Data Storage and Management* module pre-processes and stores the received data. It has the role of structuring and protecting the data appropriately. A *Human Model* exploits the data to reflect the status and behaviour of the twinned human. Lastly, *Human Digital Twin Functions* may use both the data and the human model to generate the desired feedback for the twinned human. The components are described in detail later.

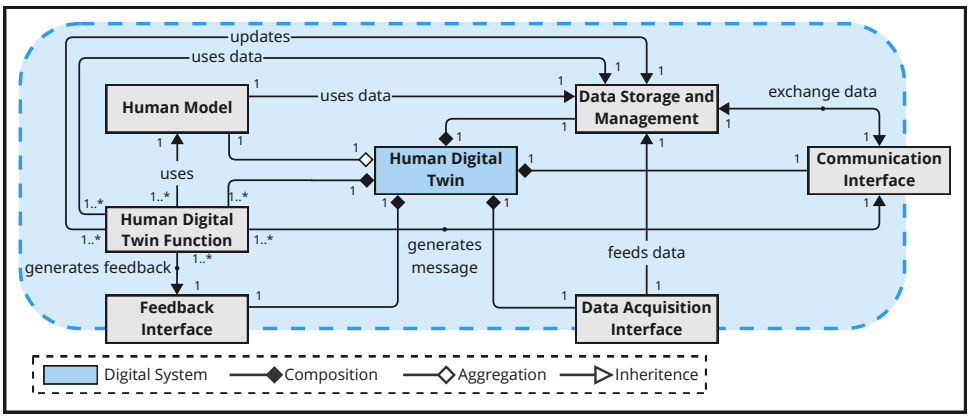


Figure 2.7: Meta-Model for Human Digital Twin.

The boundary of the HDT is materialised by its interfaces, which facilitate data exchange with external systems and the environment. A first interface is the *Data Acquisition Interface* whose role is to connect data sources about the twinned individual to the HDT. It can only receive digital signals; therefore, the sensors need to transform the detected physical signals into digital signals before transmitting them. The *Data Acquisition Interface* is mandatory to enable the necessary automatic data gathering for the HDT. Feedback to the human is managed by a *Feedback Interface*.

Lastly, a *Communication Interface* has the task of connecting the HDT with other DTs or technical systems, which can be considered the digital environment of the HDT. It is a key component to enable adaptive systems and enriching the HDT with contextual data. Through it, contextual information about the environment can be fed into the HDT and exploited for context-aware models. At the same time, it allows the HDT to send information to other technical systems for adaptation. The *Communication Interface* is a necessary component to integrate an HDT with a larger digital ecosystem, such as a smart factory.

Figure 2.7 illustrates the HDT as a *Digital System* indicated by the blue box. The interfaces are depicted on its boundary to indicate their role as a bridge between the HDT's internal components and the environment. When data is received through the *Data Acquisition Interface*, they are sent to the *Data Storage and Management* module. This module cleans and pre-processes the data before storing it for further use. It may receive additional data on the environment through the *Communication Interface*. Organising data from heterogeneous sources and fusing the data appropriately is a considerable challenge [Pesapane et al., 2022]. For real-time applications, data streams may also be set up. At the level of the *Data Storage and Management* module, access policies and privileges are managed, thus enabling the protection of personal data and privacy. What kind of data are available for sharing needs to be decided here.

The *Human Model* in Figure 2.8 is at the core of the HDT in its role of emulating the twinned human. A successful human model produces the same output as the real system for a given input. It uses data from the *Data Storage and Management* module to update itself and reflect the state and behaviour of the human. The human model may be composed of multiple specialised models, each addressing a different aspect of the human while being interlinked with one another. In practice, the scope of the model is restricted to a few aspects, as the full complexity of the human exceeds current modelling abilities. The model outputs are a potential input for the *Human Digital Twin Functions*.

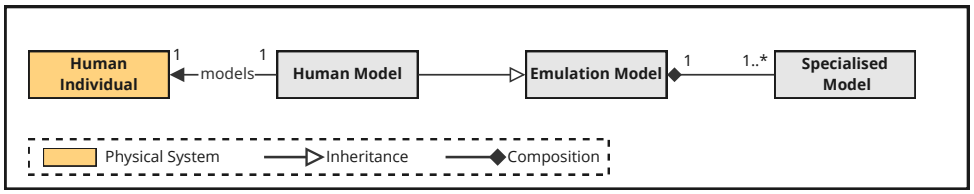


Figure 2.8: Human model: the Emulation Model of the Human Individual, as a composition of specialised models of the HDT's twinned entity.

Human Digital Twin Functions can take various forms with the goal of generating feedback to support the human. They may use the data and models of the HDT and can trigger automatically or on demand by the human. A noteworthy example is the simulation function for which the model is run with hypothetical input data to analyse the behaviour of the human under those circumstances. To run such a simulation, the circumstances and their evolution need to be modelled as well. HDTs can also be used to predict the state of a human in the near future by running the human model for multiple time steps. This may be used to avoid dangerous situations by sending pre-emptive warnings. The manually triggered function may include data analysis. For instance, an athlete may query their HDT's analytics function to visualise the trend of their fitness. A more detailed discussion on the relation between the human model and HDT functions is provided in section 2.3.7.

2.3.4 Human Meta-Model

A specific challenge for HDTs is creating fit-for-purpose models of the twinned human and use them for emulation purposes. In this Section, we present a meta-model for humans that outlines and organises the main components needed to emulate human behaviour and actions.

We conceptualise the *Human Individual* as a *Physical System* that consists of a set of components or sub-systems, and a boundary to separate it from the environment. The boundary would include the physical body of the human, which is best delineated by the skin. All internal organs, muscles, nerve strains, etc., are components of the human body and give rise to a large variety of processes, including biomechanical and cognitive processes. Properly defining the interfaces with the environment is challenging and depends on the modelling scope and perspective. For example, a diet model may consider the mouth and excretory system as interfaces, while a cognitive model would consider sensory and motor organs as the main interface with the environment.

We consider that the human can be divided into a set of sub-systems, enabling a more focused modelling approach. In Figure 2.9, we illustrate the division into *Physiological*, *Mechanical* and *Cognitive* sub-system with a detailed view on the *Cognitive System*. Our human model is composed of the three systems, but additional systems may exist. Each of the systems may be further divided, for example, the *Physiological System* may be subdivided into a cardiovascular, respiratory, and immune system. Each sub-system has attributes, but the interplay between multiple sub-systems may lead to the emergence of additional global attributes. For example, both the cognitive and physiological systems may jointly impact the psychological attributes of a human.

Figure 2.9 gives a closer look at the interactions between the *Human Individual* and *Environment* take place. It involves two major systems; the *Mechanical System* to perceive the stimuli from the Environment and execute *Physical Actions*; and the *Cognitive System* to process stimuli and select what actions to perform. Here, the interfaces with the physical world are the *Sensory*

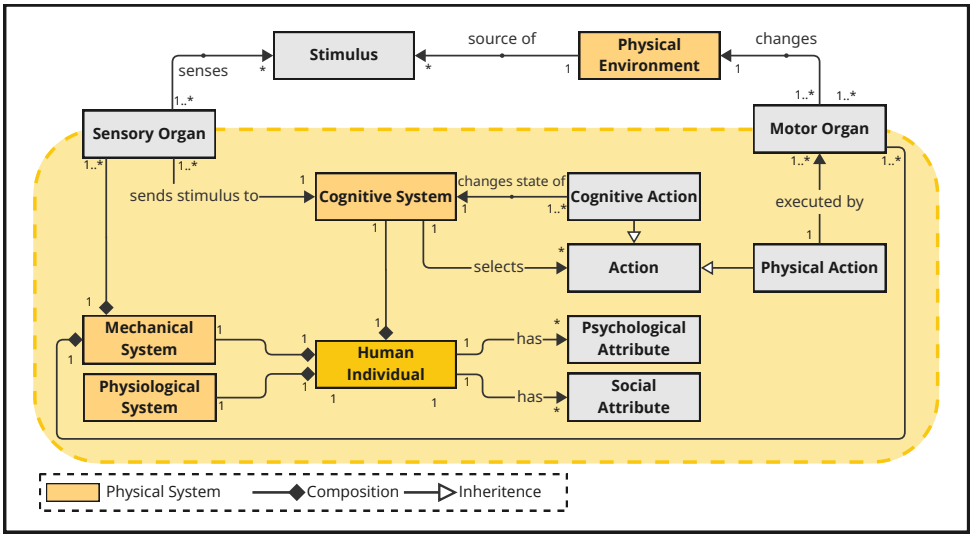


Figure 2.9: Meta-Model of a *Human Individual* as a *Physical System*, including the main subsystems and interaction with the *Environment*.

Organs and *Motor Organs* which both belong to the Mechanical System of the human. The environment is the source of a stimulus that is sensed by the sensory organs. Subsequently, the cognitive system processes this stimulus to select an action and create a motor plan. Through the execution of the action, the environment changes, generating a shift in the sensed stimuli or a new one. A human may also take a *Cognitive Action* as a reaction to a stimulus that leads to a state change in the *Cognitive System*. An example is a human mentally replanning how to perform a task after noticing faulty equipment. In this model we conceptualise the *Cognitive System* as a *Physical System*, as it can be understood as being the physical brain that processes stimuli. However, it needs to be recognised that the nature of cognition is a long running philosophical debate dating back to the Cartesian mind-body separation. For our model we take a physicalist point of view [Kim, 2007] without making any definitive statements on the philosophical debate.

Modelling human cognition is a major challenge, but a promising cognitive modelling approach comes from the cognitive sciences. These models are known as cognitive architectures and seek to model human cognition based on the assumption that the human brain works as an information processor. Their structure is informed by insights about the brain regions and their functions. Section 3.4.2 and section 3.4.3 present an exploration of cognitive modelling approaches and cognitive architectures.

2.3.5 Sensors & HMI Devices

In this Section, we describe two systems required to establish automatic bi-directional communication between the HDT and the Human Individual, namely *Sensors* and an *HMI Devices*, and discuss the data flows they enable. Finally, we consolidate all the systems discussed thus far into a holistic systemic meta-model.

We use the vocabulary of the SOSA/SSN ontology [World Wide Web Consortium, 2017] to describe the *Sensors* as part of our meta-model. *Sensors* detect observable properties through stimuli. In our meta-model we use the shorthand *observes* to describe a sensor detecting a stimulus representing an observable property. *Sensors* are composed of physical components to detect physical signals, as well as digital components to store, transmit, and potentially

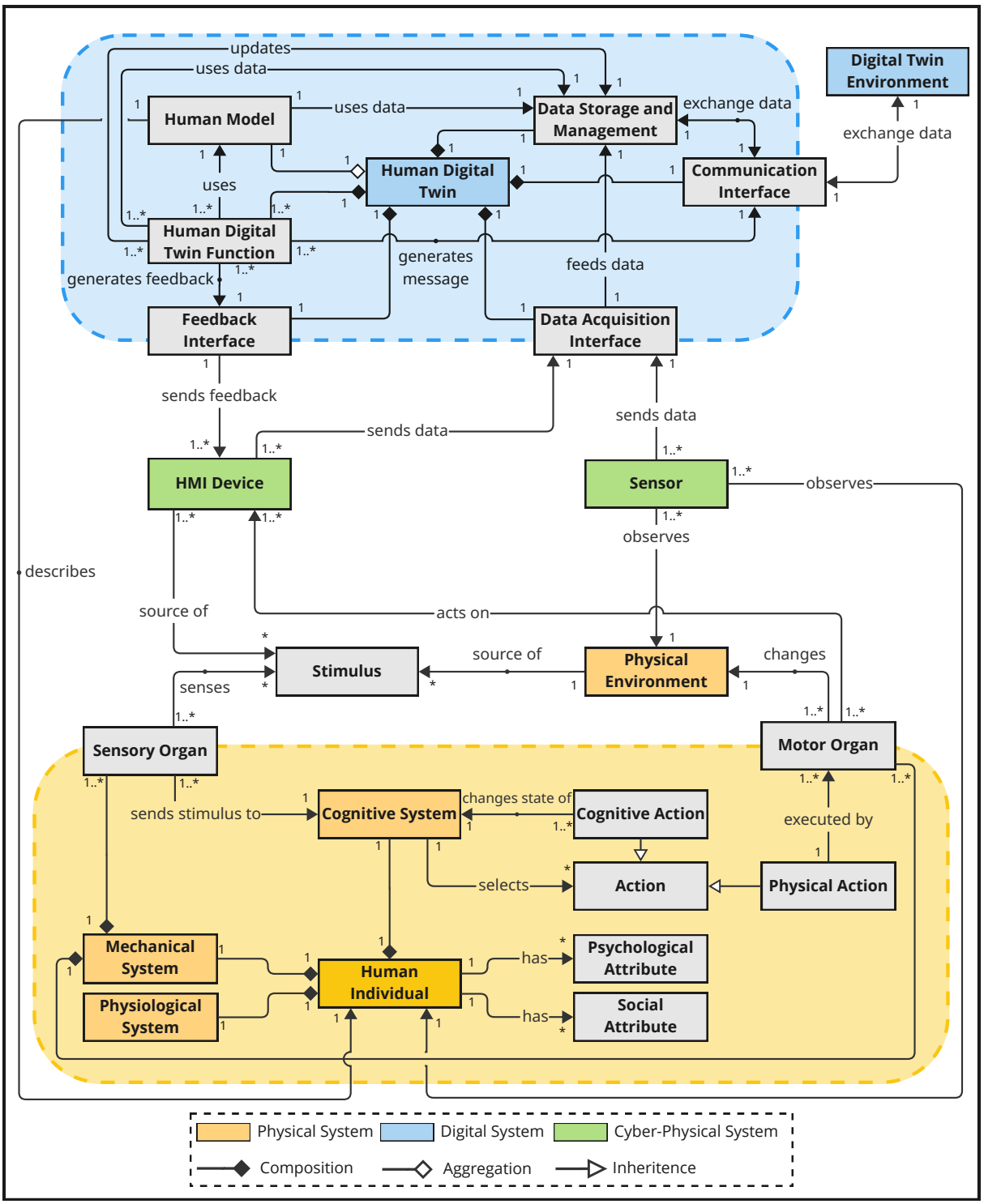


Figure 2.10: Consolidated meta-model for the *Human Digital Twin*, *Human Individual* including the information transfers by both *Sensors* and *HMI Devices*

pre-process the physical information. The exact boundary between the two is debatable, but for this thesis we define the boundary at the conversion from a physical to a digital signal, or vice-versa. An example of this is a microphone. It is composed of physical components (membrane, amplifier, housing, analogue-to-digital converter) and digital components (signal processor, transfer protocol) with the analogue-to-digital converter at the boundary.

The *Sensors* observe physical phenomena originating from physical systems, including the human individual and its environment, and transmit the information to the HDT. Observing a human includes observing their actions and the effect on the environment. From the state and evolution of the environment, it is possible to capture information about the stimulus that the human receives.

The *HMI Devices* have two roles: transforming digital signals from the HDT into human-understandable physical signals and transforming manual physical entries by the human into digital signals for the HDT. Like for *Sensors*, it has both physical and digital components to interface with the human and HDT. It enables automatic feedback to be sent to the human. As an example, the HDT may suggest to the human to take a break, by sending a digital message to the *HMI Devices* that is translated into a text message on a screen. In this example, the screen is a physical component of the *HMI Device* while the digital processes to receive and transform the HDT's message are digital components. Inversely, a physical signal, such as a keystroke, can be translated into a digital signal and sent to the HDT.

2.3.6 Consolidated Systemic Meta-Model

Figure 2.10 shows the consolidated meta-model including how *Sensors* and *HMI Devices* enable bi-directional data flow between human and HDT. *Sensors* observe both the *Human Individual* and *Environment*. This includes the physical actions that the human performs and their impact on the *Environment*. The produced data is sent to the HDT for emulation and feedback generation. Feedback is sent through *HMI Devices* which translates it into a physical signal, that manifests as a stimulus for the human which impacts their behaviour. The human may also choose to perform a physical action to send a message to the HDT through the *HMI Devices*. Finally, the *Communication Interface* enables the HDT to receive and share data with the Digital Twin Environment, composed of the other Digital Systems such as DTs of humans and machines, and direct communication with technical systems.

In summary, the systemic meta-model establishes a unified framework that defines clear boundaries and interfaces between all interacting systems. It is designed with the goal to facilitate HDT development that can be thought of in two parallel phases; ***(i)*** the development of the human model to emulate the twinned individual and the associated HDT functions; ***(ii)*** integration data sources and HMI devices to feed the model and return the HDT functions feedback. By considering data sources and feedback mechanisms as external entities, a modular development approach can be adopted. An initial conceptualization phase should define the objective of the HDT development and derive the required HDT functions to achieve it. Afterwards the development of the required emulation model, sensing devices and interfaces can happen concurrently.

2.3.7 Human Digital Twin Functions

This section briefly discusses the relation between the HDT's human model and its functions. The full in-depth discussion was published in Naudet et al. [2024] and Al Haj Ali et al. [2025a]. Figure 2.11 illustrates the relation between HDT and the twinned human, providing a more

detailed view on data flows and HDT functions. In line with the human meta-model (see Figure 2.9), the human is conceptualised as an agent that responds to stimuli `sti` from the environment with a motor action `act`. The sensory organs sense a stimulus `sti` leading to an observation `obs`. Based on the observations and internal state, the human's cognitive system making a decision `dec` on what action to execute, which leads to an observable action `act` in physical space. The HDT receives data about the environment, stimulus, other entities and other DTs. Data about the human state `data_h` can be sensed by deploying sensors. Given the limitations of sensing capabilities, many parameters of a human's internal state remain hardly accessible. A second source of information about the human are manual data entries. The human may choose to provide information through an action `act_m` that need to be converted into exploitable data `data_m` by an HMI. Feedback `fbk_data` from the HDT cannot be directly received in the form of data. Instead, it is necessarily transformed by the HMI into a physical stimulus `fbk_sti` that can be perceived and understood by the human. The HDT has three major functions; (1) *Emulation*, (2) *(Cognitive) Functions*, and (3) *Simulation*.

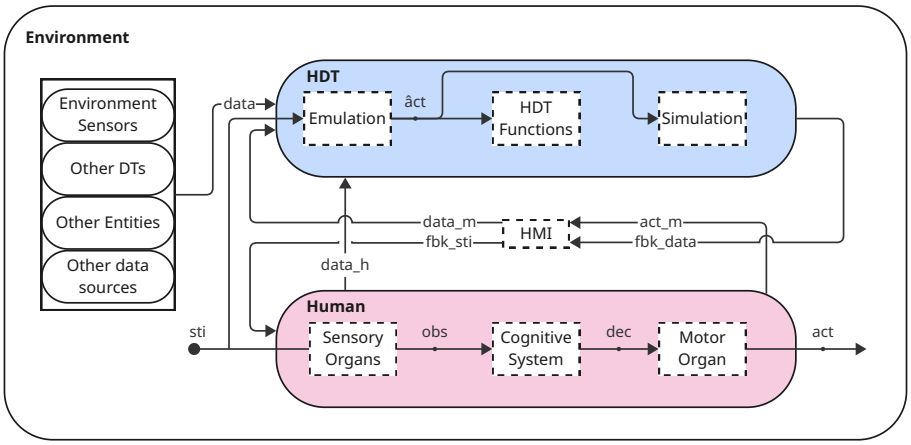


Figure 2.11: Human-HDT architecture with its main functions.

Emulation is the central function of the HDT. It computes an estimation of the twinned human's response `âct` to a stimulus `sti`. We formalise the emulation models of cognition in section 3.4.1. At this stage, we only describe the input-output relations and how the resulting output can be used by other HDT functions. Emulation is dependent on the measurable data about the human `data_h` and the received stimuli `sti`. Both need to be sensed with sensors as described in the consolidated meta-model in Figure 2.10. The emulation function estimation `âct` can be validated against the observed actions `act`, and the error between estimation and observation $||act - \hat{act}||$ minimised [Naudet et al., 2024].

HDT functions: Based on the emulated human action and sensed data a variety of functions may be implemented. Discrepancies between estimated actions from the emulation model and observed actions can be interpreted in two ways; (1) the emulation model is inaccurate and needs to be calibrated or trained to more closely follow the real human behaviour; (2) the human is exhibiting abnormal behaviour for which an intervention may be needed. In the first case, the new observations may be used in subsequent model training (if data-driven models are deployed). For parametrised mechanistic models it is an indication that a calibration step is required. In the second case, the emulated behaviour can be considered the one of an undisturbed and healthy

human being. Under that assumption discrepancies can be an indication of an abnormal state or dangerous behaviour. For example, the human may have a medical emergency which exceeds the modelling scope of the emulation model. A less dramatic example is a worker that is distracted by external factors that are not captured by the HDT (e.g. a loud noise when the HDT is unable to receive auditory data). If an intervention is needed the HDT can do so by sending a feedback `fbk_data` to the HMI which is transformed for the human into a feedback stimulus `fbk_sti`, for instance a message on a screen or an alarm sound. Alternatively, the HDT can send messages to external system to adapt to the unexpected human behaviour. For high fidelity emulation models, the estimated action of the emulation model can be used to provide proactive guidance. For instance, an estimated action can be interrupted if it creates a risk for the human or may cause damage to equipment or components. HDT functions can also directly exploit the tracked state of its twin. A simple example would be a function that triggers a task reallocation when the tracked fatigue level exceeds a predefined threshold.

Simulation function: The HDT's simulation function can be used to explore new situations including the possible reactions of human workers. Simulations with HDTs require a simulator of the environment to resolve the simulated actions and generate the new stimuli. Exploring new situations is particularly useful if multiple DTs are available to co-simulate new situations. For instance, simulations of new workspace configurations could be used to assess their impact on ergonomics and production efficiency. An additional use of the simulation function is evaluating the probable human reaction to feedback. It would help avoid sending feedback that risks interrupting the twinned worker during critical moments that may have a negative effect on production outcomes or safety. In general, it may help provide relevant information at right time and in the right format to guide the worker without causing unnecessary cognitive load.

2.4 Conclusions

In this Chapter, we propose a general definition for Human Digital Twins, a categorization system based on the level of data integration, and a meta-model grounded in general systems theory. The proposed contributions synthesize the fundamental characteristics of HDTs based on the wide body of literature surveyed in Chapter 1. The general definition identifies the required elements for HDTs, while the systemic meta-model defines their roles, boundaries and interfaces. Jointly the two contributions answer the first research question of this thesis **RQ1**. The general definition and categorization system address the prevalent conceptual ambiguity in HDT literature (**Limitation 1**), while the systemic meta-model addresses the lack of a unified meta-model (**Limitation 2**).

It is important to note that the proposed meta-model represents a simplified approximation of the twinned human. While this perspective relies heavily on automatically detectable signals, such measurements capture only a subset of the information relevant to a human individual. Indeed, human behaviour is shaped by additional factors that are not directly detectable with sensors. Notable examples are biographic information, social characteristics, personality traits or acquired abilities [Naudet et al., 2021b], which are undetectable via direct sensing but can be established through questionnaires. The vast majority of human complexity remains largely inaccessible to current sensing technology. Due to this limitation, the challenge becomes identifying internal states of the twinned entity based on limited external observations. Additionally, the parametrisation of models to emulate a human individual remains a complex task. Therefore, it is advisable to think of HDTs as a fit-for-purpose approximation of the twinned individual rather

than an exhaustive and complete digital double.

Creating true HDTs requires not only the creation of an emulation model but also its integration. The challenge of creating an emulation model, especially for cognitive aspects, remains among the hardest challenges. As our aim is to enable adaptations in HRC setups based on modelled human behaviour, we will focus the entire remainder of the thesis on cognitive modelling and its exploitation. Chapter 3 proposes an architecture to enable adaptations, while Chapter 4 homes in on the challenge of creating an emulation model for human cognition.

Chapter 3

Human Digital Twins for Human-CPS Collaboration

3.1 Introduction

The first two chapters of this thesis established the theoretical foundations of HDTs through a general definition, a classification of HDTs, and systemic meta-model. Building upon these foundations, this third chapter investigates the potential use of HDTs to achieve the core objectives of the AI4C2PS project; extending interoperability from the semantic to the cognitive level between humans and CPS (see Section 1.1). As established in the project's framing, AI4C2PS addresses the limitation that humans and machines process knowledge in fundamentally different ways, necessitating a move from semantic to cognitive interoperability. The project adopts a decentralised strategy of digital twinning where each agent, whether human or CPS, is paired with its own DT as illustrated on Figure 3.1. For CPS the DT is a Cognitive Digital Twin (CDT) [Al Haj Ali et al., 2024a] which has the objective to provide the engineered system with the cognitive abilities required to learn and adapt in collaborative industrial environments. For humans, the HDT is positioned as the key technology to emulate human cognitive processes at runtime, providing personalised support and **s**haring information about the human's cognitive state with collaborating entities. Both DTs can communicate with each other to make joint decisions and orchestrate the behaviour of their twin (for CPS), or provide guidance and recommendations to their twin (for humans).

For our investigation, we utilize a use case centered on Human-Robot Collaboration (HRC), an industrial domain where awareness of the human state is critical for ensuring safe and effective interaction. First, we present contemporary approaches to HRC including the emerging use of HDTs to adapt to human states. Of particular interest are studies where the human's cognitive state is tracked and modelled for adaptive robotics. To achieve the AI4C2PS' stated goal of cognitive interoperability, we focus this chapter's scope on cognitive and behavioural aspects with the objective to answer research question (**RQ2**) *How can HDTs be used to track and emulate human cognition and behaviour? In particular, how can this improve collaboration between humans and Cyber-Physical Systems?*

This leads to two sub-questions: ***(i)*** How can an HDT that emulates human cognition be exploited for CPS adaptations? ***(ii)*** How can human cognition be modelled and emulated? First, we discuss the merits and challenges of using HDT-based approaches, detail the data flows in a high-level architecture, and illustrate that approach with a Human-Robot Collaboration scenario. Second, we discuss the emulation model that forms the heart of the HDT. We formalize what an

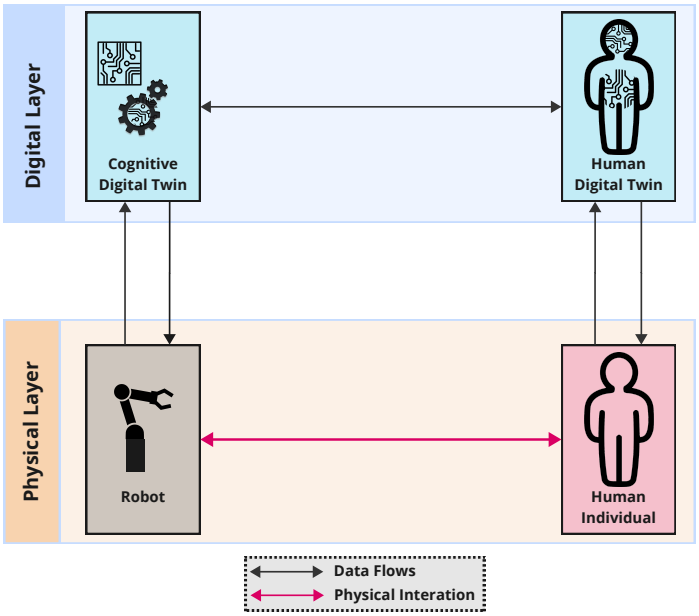


Figure 3.1: Conceptual architecture for Human-Robot collaboration from the AI4C2PS project.

emulation model for cognition is, including its interaction with the environment. Subsequently, we explore existing methods, architectures and models from cognitive modelling to find the best candidate to create an emulation model.

This chapter draws from multiple publications made as part of this thesis.

- Gaffinet et al. [2023] presents challenges in Human Robot Collaboration and an initial draft of the DT-based architecture. The overview on HRC challenges is partially from this publication.
- Al Haj Ali et al. [2025b] provides the definition for cognitive interoperability that serves as context for the chapter.
- Gaffinet et al. [2025b] presents the refined DT-based architecture, and possibility to use cognitive architectures, specifically ACT-R, to emulate human cognition.
- Bhattacharya et al. [2026] contains common elements with the emulation model formalization presented in this chapter.

3.2 Challenges in Human-Robot Collaboration

Interactions between humans and robots have evolved over time, starting with mere coexistence in the same physical space with completely independent operations. Eventually the interactions between humans and machines deepened. Wang et al. [2019] distinguishes Coexistence, Interaction, Cooperation and Collaboration between Humans and Robots. Among these, collaboration is the most advanced mode of interaction. It is defined by a shared physical workspace, a shared common task, shared resources, simultaneous actions and finally physical contact is allowed between human and robot. More advanced relations have been discussed, for instance the concept

of co-evolution which is defined as a relation where humans learn from machines and vice-verse [Lu et al., 2022]. However, for this chapter we restrain our scope to HRC, as its challenges are well established and provide fertile ground to explore the potential of HDT as an enabling technology.

3.2.1 General HRC Challenges

In collaborative setups, the close contact between robots and humans comes a non-negligible risk of injury which motivates extensive research into robot safety methods and collision avoidance in particular. Real-time situational awareness is a common challenge and required to estimate distances or distinguish voluntary from accidental contact. Robots need to be more flexible and adaptable in their programming to work with the uncertainty that human behaviour introduces into the system. Nonetheless successful HRC implementations can improve ergonomics [Lorenzini et al., 2023] by minimizing mental and physical fatigue by assigning repetitive and heavy lifting tasks to the robot. But poorly implemented systems can induce stress in the human worker and erode trust in the machine. Wang et al. [2019] identify essential properties of HRC and stress that: *"An essential aspect of HRC is how to cope with human ergonomics, process time, emotions and reaction during collaboration, and safety aspects."*.

Krüger et al. [2009] discuss the merits of HRC, making the case for how robot reliability and human adaptability can enable a new generation of assembly processes. However, safety becomes a prime concern as soon as the physical separation between human workers and industrial robots is no longer guaranteed. HRC, by definition, allows physical contact making physical separation inherently impossible. The causes for HRC accidents are organized in three categories; engineering failures, human errors, and poor environmental conditions [Vasic and Billard, 2013; Wang et al., 2019]. For robotics and machinery in general, engineering failures are addressed in the design phase, and guidance is provided in ISO 13855 [ISO13855, 2010]. The unavoidable physical contact between robots and humans in HRC setups required an adaptation of standards, which lead to ISO/PDTS 15066 [ISO/TS 15066, 2016]. It provides guidance on risk assessment, safety features, workspace design as well as task and process design. Overall safety in robotics is a well-researched subject, nonetheless the challenging context of HRC requires further investigation on how to enable setups that can operate at relatively high efficiency and speed without compromising safety.

Research in HRC safety can be categorized into three categories; (i) understanding injury mechanisms and related standards; (ii) limiting the severity of an impact; (iii) active collision avoidance. In the first category, Haddadin et al. [2008b, 2007, 2009] made major contributions with their work on injury mechanisms of collisions that might occur during HRC task execution. Further research establishes that safety is guaranteed for robot velocities below 2.7 m/s [Haddadin et al., 2008a] and the link between mass, velocity, impact geometry and resulting injury [Haddadin et al., 2012]. Instead of limiting operational velocity, Laffranchi et al. [2009] proposed to regulate the energy of the system instead. Earlier research from Yamada et al. [1996] analyses pain tolerance of humans and establishes an early method to reduce robot velocity upon impact, which allows the human operator to reflexively withdraw and avoid serious injuries and pain. More recent research introduces additional constraints to the control algorithms based on newly introduced safety indicators [Meguenani et al., 2016]. Depending on the distance between robot and human operator the kinetic and potential energy in the system is reduced to limit the energy that could be dissipated in the event of a collision.

Detecting collisions and reacting appropriately represents a second major category of research in HRC safety. Unmitigated movement of a robot after an unwanted collision can lead to severe

pain and injury. As contact between robot and human is expected in HRC it is not appropriate to stop operations as soon as any physical contact occurs, instead more subtle methods are required. On the hardware level the addition of joint torque sensors enabled a new set of safety features. Tonietti et al. [2005] adjusts the stiffness of actuation to inherently guarantee safety. Park et al. [2008] developed a passive safety mechanism through non-linear stiffness. During normal operations high stiffness is maintained, but upon impact, that exceeds a set threshold force, the stiffness drops quickly.

Avoiding collisions altogether is approached from different angles by researchers. A large swath of research uses a combination of detection systems such as, time of flight cameras (or other depth sensors) [Ahmad and Plapper, 2016; Anton et al., 2012; Flacco et al., 2015] or optical cameras [Tan and Arai, 2011]. Creating 3D models of the robot, environment and/or human worker is sometimes used in addition to the chosen sensing systems to enable better distance estimations [Mohammed et al., 2017; Wang et al., 2013]. Some approaches put the human worker at the center of their investigation by tracking their movement, either with wearables or external cameras to ultimately include human behaviour [Zhao and Pan, 2018] as a parameter for robot adaptation [Liu et al., 2018]. Decades of Robotics safety research have led to hardware adaptions, and methods that substantially reduce the risk of severe injuries. Avoiding collisions altogether remains a challenge and is mostly done by reacting to detected human movements.

A second major line of research for HRC is ergonomics. Robots should support the human worker to fulfil a defined task as a human-robot team while maintaining their well-being [Wang et al., 2019]. Ergonomics encompasses not only physical stress, which can be induced by e.g. lifting heavy loads, but also mental stress. A necessary component of HRC is that the human team member accepts to work with the robot. Poorly adapted robot movement can induce stress, erode trust and ultimately lead the loss of propensity to work with the robot. Trust is a major parameter for successful HRC setups and has been studied extensively by Hancock et al. [2011]. The dominant factor that can erode trust is the performance of the human-robot team, while environmental and human factors have a more limited impact. The distance between human and robot alongside the robot's movement speed has a direct impact on induced stress [Arai et al., 2010]. The more general impact of robot movement on the human affective state has been investigated by Kulic and Croft [2007]. Adapting the robot's movement by designing human-aware systems is a promising approach that has found success in the field. Research addressed the problems through better motion planning based on human-aware systems. The movement needs to be predictable and perceived as safe. Multiple paper seek to improve motion planning and task allocation by including information about the human worker; such as the human's goals [Pellegrinelli et al., 2016], mental model [Nikolaidis and Shah, 2012], perception of robot movement [Dragan and Srinivasa, 2014; Pandey and Alami, 2010], next subtask [Hawkins et al., 2013] or predicted motion [Unhelkar et al., 2018]. Many promising approaches to improve fluency of HRC systems and ergonomics rely on understanding the human worker's behaviour and state better which is a promising application area for HDTs.

3.2.2 Human Digital Twins in HRC

HDTs are viewed as a promising approach to some of the contemporary challenges in HRC that extend beyond traditional safety standards. As highlighted by Asif et al. [2026], HRC setups must address not only physical safety but also perceived physical safety, as the mere presence of robots can induce discomfort and deteriorate psychological well-being. Trust management emerges as a critical factor, where excessive trust may lead to overuse and dependence, while distrust risks the

abandonment of the technology. Additionally, human idle time must be carefully managed, and HRC systems should avoid seeking to replace the human element entirely. A further limitation lies in current perception systems, which struggle to reliably recognize human intent, complicating seamless interaction. The psychological dimensions of HRC are equally significant. Carissoli et al. [2024] demonstrate that cobot motion patterns, speed, pace, and role in task allocation are associated with higher mental workload. Their work calls for greater interdisciplinary collaboration to integrate expertise on human psychological functioning. Beyond mental workload and stress indicators, repeated subjective metrics assessments remain under-explored yet are deemed important for HRC. Predicting human behaviour in HRC introduces additional complexities. Du et al. [2026] emphasize that the uniqueness of human behaviour, such as variations in body size leading to different trajectory selections in simple tasks, makes predictions inherently difficult. Individual predispositions complicate accurate predictions. Dynamic states like trust, fatigue, and workload further impact behaviour while the absence of a physical world model can result in implausible predictions. Dell'Oca et al. [2026] add that high prediction accuracy alone is insufficient for effective collaboration. Through a user study, they found that contextually appropriate and non-disruptive robot behaviour can be more critical than raw accuracy. Time dependence of actions and the interpretability (legibility) of those actions are major drivers for trust, reinforcing that interaction design and workflow are of great importance.

The challenge of deducing another agent's intentions is central to the cooperative agent model proposed by Pacaux-Lemoine et al. [2017]. This framework posits that effective collaboration requires the ability to *Know-How-to-Cooperate*, which includes understanding the intent of other agents and providing them with information. HDTs represent a promising technology for supplying the information needed to enable collaborating systems to understand their goals and intentions.

These challenges have an important communality; they are caused by the psychological predispositions and behavioural tendencies that are not easily detectable through direct sensing. Reactions and results may differ for each individual which cannot be captured with generic models. HDTs emulate individuals and are consequently well positioned to address these challenges. While the field is young, there have been notable advances in applying HDTs, or general DT-based approaches, to HRC scenarios.

Langås et al. [2026] discuss the advantages and limitations of DT based solutions in Industry 5.0. Robot DTs can aid collision avoidance, while virtual testing of new equipment and setups can reduce prototyping costs and downtime. HDTs, in particular, are seen as potential sources of behavioural and trajectory information about humans, ultimately enabling dynamic robotic adaptation. However, the predictability of humans remains a major bottleneck. Zhu and Yang [2025] review HDT applications broadly, including for HRC, and present a case study for a collaborative assembly task. Their framework, comprising one HDT and one Robot DT, identifies promising applications such as robot motion trajectory planning based on human intent for collision avoidance and dynamic task allocation based on human fatigue. In this setup, the Robot DT coordinates the collaboration, while the HDT serves purely as a source of information, acting as a decision-making support technology for operators that can adapt dynamically to changing environments. For tasks requiring close physical proximity, such as handover tasks, HDTs show promise. Zhong et al. [2025] highlight that adapting to the detailed movements of humans is a major challenge in such scenarios. Their implementation of an HDT incorporates both physiological and physical states (skeletal data) of the twinned human, enhancing the robot's dynamic perception. Intentions are detected through temporal features, and while the paper demonstrates results on intention recognition, the integration of this model for robotic adaptation

is not yet realized. Nevertheless, the use of HDTs is recognized as highly promising for optimizing assembly sequences and enabling proactive collaboration, where the robot understands human intention and can take initiative accordingly. Liu et al. [2025] further identify the potential of DTs to enhance human wellbeing in HRC setups, particularly in terms of human comfort. Real-time assessment remains a major challenge, as current methods rely heavily on subjective post-task questionnaires and interviews, which fail to provide real-time insights. The authors propose a method based on physiological state and fatigue detection in real time, which, when deployed in an HDT, can enable real-time feedback such as posture adjustments or break suggestions. Baratta et al. [2024] echo these sentiments, identifying DTs as potential tools for human-centric systems. They emphasize that DT-based setups can support collision avoidance, diagnostics, and prognostics to maintain safe work conditions. The DTs simulation function is a valuable tool to assess collaboration strategies or workcell adjustments without putting the human workers or physical assets at risk.

Despite these advancements, the integration of human data in industrial Cyber-Physical Systems (CPS) through Human Operator Digital Twins (HODT) raises significant ethical concerns. Cardin and Trentesaux [2022] emphasizes key ethical risks that include treating humans as disposable machine-like resources, violating the intimate sphere via intrusive sensing, and the standardized over-usage of the operator's natural ability to absorb temporary work overloads. There is also a critical risk of misusing data for firing decisions or discriminatory HR practices. To mitigate these risks, the source recommends disconnecting HODTs from global company networks to limit undesired data mining and avoiding fully automated scheduling. Finally, designers must apply "dampening", acknowledging that a digital twin cannot capture the actual essence of its physical counterpart, to prevent overreacting to predicted performance gaps.

In summary, HDTs have been identified as a promising technology to address key challenges in HRC, including safety, adaptability, and human well-being. HDTs are well positioned to provide real time intent information and enable collaborating robots to adapt to anticipated actions. They can track physiological and ergonomic parameters enabling real time adjustments. Workplace adjustments and collaboration strategies can be simulated for assessment, without putting physical assets and workers at risk. The risk of data misuse is substantial, even though HDTs could be designed to constrain data use to the required local context to mitigate the risk of misuse.

3.2.3 Cognitive Interoperability

The AI4C2PS project proposes extending interoperability beyond the semantic level to account for the fundamental differences in how humans and machines process information. Effective collaboration between two agents, whether human-human, human-system, or system-system, requires the ability to share and act upon information. This necessity forms the basis of interoperability, ensuring that agents can cooperate despite differences in language, structure, or internal representation. Without interoperability, collaboration is susceptible to failure due to misunderstandings, incompatible data formats, or conflicting expectations. The European Interoperability Framework (EIF) defines interoperability layers [European Commission and Directorate-General for Digital Services, 2017]:

- *Technical interoperability* refers to the ability of systems and services to exchange data. This includes secure communication protocols and interface specifications. Legacy systems can lead to fragmented landscape that complicates technical interoperability.

- *Semantic interoperability* in the EIF includes semantic and syntactic aspects. Syntactic aspects refer to a clear definition of formats and grammar used for data exchanges. The semantic aspect assures that the meaning of the data is understood by both parties. Ontologies are a widely used technology to formally encode the meaning and relation of exchanged data and avoid any misunderstanding about their meaning.

- *Organisational interoperability* refers to the alignment of business process and definition of responsibilities and expectations. In practice this involves agreeing on and documenting business processes to achieve mutually beneficial goals.

- *Legal interoperability* is the process of ensuring that organizations operating under different legal frameworks, policies, and strategies are able to work together effectively.

Higher level interoperability levels namely, organisational and legal interoperability, do not directly impact the quality if HRC. Semantic interoperability is designed to enable machine to machine communication. Due to this original purpose, it does not address the different ways humans and machines process data. It does not address how humans process and understand data. This is shaped by internal cognitive processes that depend not only on the formal meaning of the data, but also on the current context, previous experiences, personal goals, and expectations. For instance, two individuals may receive the same information, yet interpret it differently based on their background and current situations. Semantic interoperability, which focuses on shared definitions and vocabularies, cannot capture these differences in individual perception and reasoning.

Furthermore, technical, syntactic, and semantic interoperability are all made possible by developers and engineers defining how data is exchanged, formatted, and interpreted. The entire process, from receiving data to making decisions, is governed by logic and rules designed in advance. This makes the system's behaviour predictable and controllable, with each processing step traceable. In principle, one can inspect how a decision was made and understand the reasoning behind it. However, this design assumes that all situations can be anticipated in advance and that all the knowledge needed for decision-making can be encoded at the design stage. In real-world collaboration, especially when humans are involved, this is rarely the case. Systems must be able to handle new situations, adapt their behaviour, and understand the perspectives and intentions of other agents. These requirements exceed the scope of semantic interoperability and underscore the need for cognitive interoperability Al Haj Ali et al. [2024a] which "*refers to the ability of different agents (human or artificial) to align their thoughts and perception of information, allowing for mutual understanding and shared intentions. It involves creating shared mental models, aligning knowledge, and ensuring a common way to use, interpret and reason over knowledge*" Al Haj Ali et al. [2025b].

Cognitive interoperability presents several challenges that go beyond standard data exchange. It requires agents to develop mutual understanding, align their perceptions and thought processes, share intentions, and participate in joint decision-making. Mutual understanding involves each agent being able to comprehend how others interpret a situation, taking into account their beliefs, goals, and experiences. Aligning perception and thought processes means that agents must interpret their environment in comparable ways and reason over information using compatible internal models. When this alignment fails, collaboration breaks down due to inconsistent interpretations. Shared intentions mean that agents must not only be aware of each other's goals, but also commit to common objectives, which is important to achieve coordinated behaviour. Finally, joint decision-making requires agents to act collectively based on this shared understanding and alignment of intentions. These requirements are especially important in human–system

collaboration, where the system must adapt to the human perspective and reasoning. In order to enable such adaptations, a persistent digital representation of not only the physical state, but also the cognitive state of the human is necessary. This motivates the use of an HDT, which serves as a digital counterpart to the human involved in the collaboration. To enable cognitive interoperability in operational settings, information about the internal cognitive processes and states of collaborating entities should be modelled and shared among them. Our hypothesis is that an HDT emulating human cognitive processes should be able to support this.

In summary, multiple current challenges in HRC and cognitive interoperability can be traced to a lack of awareness of the human state. Notable examples are; (i) lack of knowledge about human intent, leading to a risk of choosing conflicting actions and collision; (ii) lack of information about the human's knowledge levels about the task and setup; (iii) a general lack of a theory of mind about the human's cognitive state and hence an absence of cognitive interoperability; (iv) no real-time information about fatigue levels and the related increased risk of human errors. In the next section we investigate how HDTs emulating human cognition may mitigate the problems listed above as part of a DT-based architecture.

3.3 Digital Twin Based Architecture

We focus on the problems that involve the cognitive state and processes of the human, with the goal of improving the collaboration between humans and CPS, as well as contributing to cognitive interoperability. One interesting property of human cognition are the interpersonal differences that depend on an individual's personal experience and cognitive abilities. Given the exact same situation, each individual may still react completely differently due to their past. Capturing these interpersonal differences becomes an essential requirement for successful modelling. One may argue that a robot should be equipped with sensors to observe the human and infer the cognitive state. However, an observation made while collaborating cannot inform the robot about the human's past and personal experiences that may impact behaviour. Prior knowledge about tasks or memory capacity of an individual are also no directly observable but certainly impact the behaviour during collaboration. Lastly, also mere personal preference is not observable, but includes relevant information like an individual's tolerance for a robot's proximity.

Capturing individual differences that are undetectable with sensors mounted on a robot is the main motivation for our HDT-based approach. As established in Chapter 2, an HDT *"[...] is a digital representation of the twinned human, emulating their state and dynamics. [...]"*, where emulation means that the HDT produces the same behaviour as the twinned individual when presented with the same situation. An HDT centralizes information about their twin, which is not restricted to the observation window during collaboration, but includes personal preferences and past experiences. This richer background enables the HDT to reflect individual differences, making it a promising approach to address the previously discussed challenges.

The long-term vision sees an HDT as a persistent digital companion throughout a person's life. Where the human remains in control and the HDT has the responsibility to protect the data it holds. Within various contexts of life, the HDT connects to other entities to provide the required information to adapt to the individual's needs or provides direct guidance to its twin. However, in manufacturing a popular approach is to create a DT for an entire shop floor including the workers. But a single DT for multiple entities is incompatible with the long-term vision, because it cannot accompany the individual worker neither is it personalised. Therefore, we propose the creation of one DT per entity that can connect when a collaborative task begins.

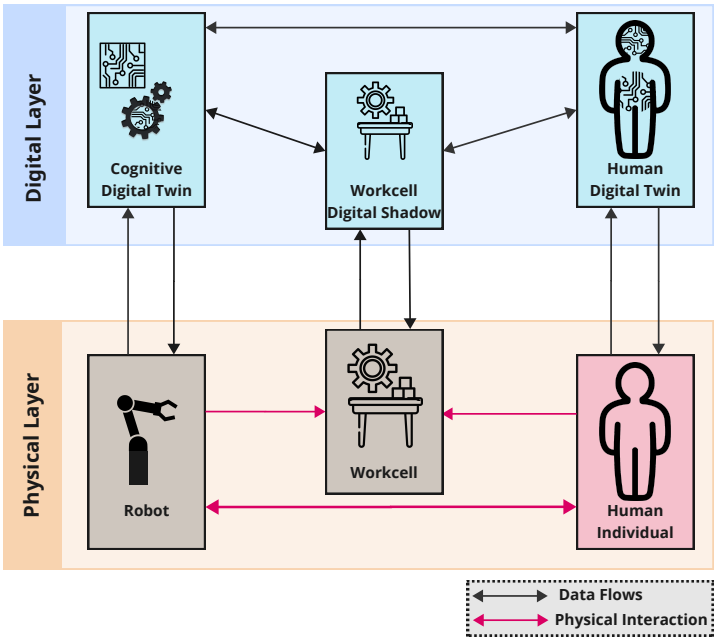


Figure 3.2: Simplified illustration of the entities in an HRC setup and their Digital Twins or Digital Shadows respectively.

Figure 3.2 shows the entities and DTs for HRC setups: ***(i)*** an HDT for the human worker, ***(ii)*** a DT for the collaborative robot, and ***(iii)*** a Digital Shadow for the workcell, tracking the state of the product, task and resources.

3.3.1 Entities and Data Flows

Consider a HRC setup with one human and one robot collaborating at a shared workstation referred to as workcell. The workcell contains all the tools and resources to perform an assembly task. In this scenario, the physical entities are the `Human`, the `Robot` and the `Workcell`. The two agents, `Human` and `Robot`, make decisions about what action to take independently to advance the shared production goal. This can be framed through a multi agent systems (MAS) lens. Selected actions can be independent from each other or interfere with each other. Figure 3.3 shows the possible types of actions in a MAS [Weyns and Holvoet, 2004]. An agent can take *independent actions* where the state of the environment changes without having any impact on other agents. Actions are classified as *interfering actions* when two agents choose actions that influence each other. Three types of interfering actions are distinguished; *joint actions* require two agents to collaborate to successfully perform the action; *influencing actions* can either positively or negatively influence each other, for example two agents pushing an object in the same direction is a positive influence while pushing in opposite directions is a negative influence; lastly, *concurrent actions* are conflicting in nature, such as two agents choosing to use the same object. The risk of collision is elevated during *concurrent actions*, which is one of the identified challenges of HRC. To avoid those actions, we propose a DT-based architecture.

Each of the three entities has their own DT (see Figure 3.2) that centralizes information about their twin and share information between each other as the task requires. First, we describe each

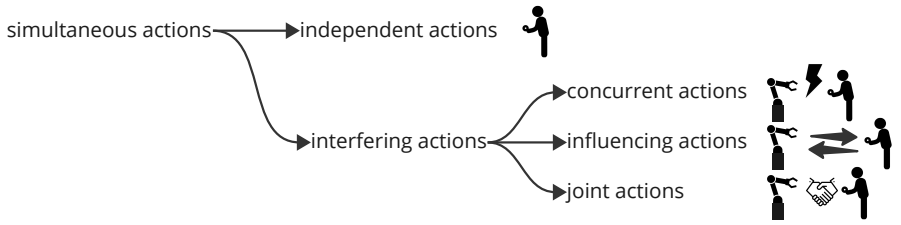


Figure 3.3: Classification of actions in multi-agent systems based on the formalisation presented in [Weyns and Holvoet, 2004]

entity and its DT before addressing the additional data flows between DTs.

Human and their Human Digital Twin:

The conceptualisation of the human and HDT follows the structure of the systemic meta-model proposed in section 2.3. For the sake of clarity, not all the subsystems of the human and not all the modules of the HDT are depicted on Figure 3.4. As our focus is on the data flows and the cognitive aspects of the human, we only include interfaces and essential components for cognitive aspects. Note that the details about sensors and HMI Devices are omitted from the Figure.

We assume that the worker has been briefed about the task, understands it, and has the goal to fulfil it. They interact with the environment by perceiving its state through their *Sensory Organs* and executing actions through their *Motor Organs*. Stimuli come from the physical environment, including the workcell, collaborating robot and messages from any HMI devices. The processing of the stimuli and selection of the next action occurs in the human's *Cognitive System* which is their brain. The selected physical action is executed and impacts the state of the workcell or robot. This processing pipeline from stimuli to action is referred to as cognitive cycle.

The HDT has the objective to emulate the human behaviour and provide guidance to its twin or information to collaborating systems. In line with the systemic meta-model, it has a *Data Acquisition Interface* (DAI) to receive information about its twin, and a *Feedback Interface* (FI) to provide feedback. Through the *Communication Interface* (CI) the HDT receives task relevant contextual information about the environment, robot and production goal. The *Human Model* is an emulation model for human cognition in this case. It models the cognitive state of the human through its *Internal State*, which includes the current goal, knowledge about the task, and awareness of the current situation. The emulation model receives information about the sensed stimuli to estimate the human's intention and action. In the process it updates the internal state such as the goal and working memory contents. *HDT Functions* have access to the emulation models output, its internal state and contextual information from other DTs. In this rich data environment, *HDT functions* can be created to generate guidance messages for the twinned human, or share select task-relevant information with the robot's CDT.

Workcell and its Digital Shadow:

The *Workcell* contains all the resources to fulfil the task. The state of the workcell can be described by the availability and position of all the components, tools and products it contains. Both the human and robot execute actions that change the state of the cell, advancing towards the production goal. The Workcell Digital Shadow (WDS) tracks all changes and communicates

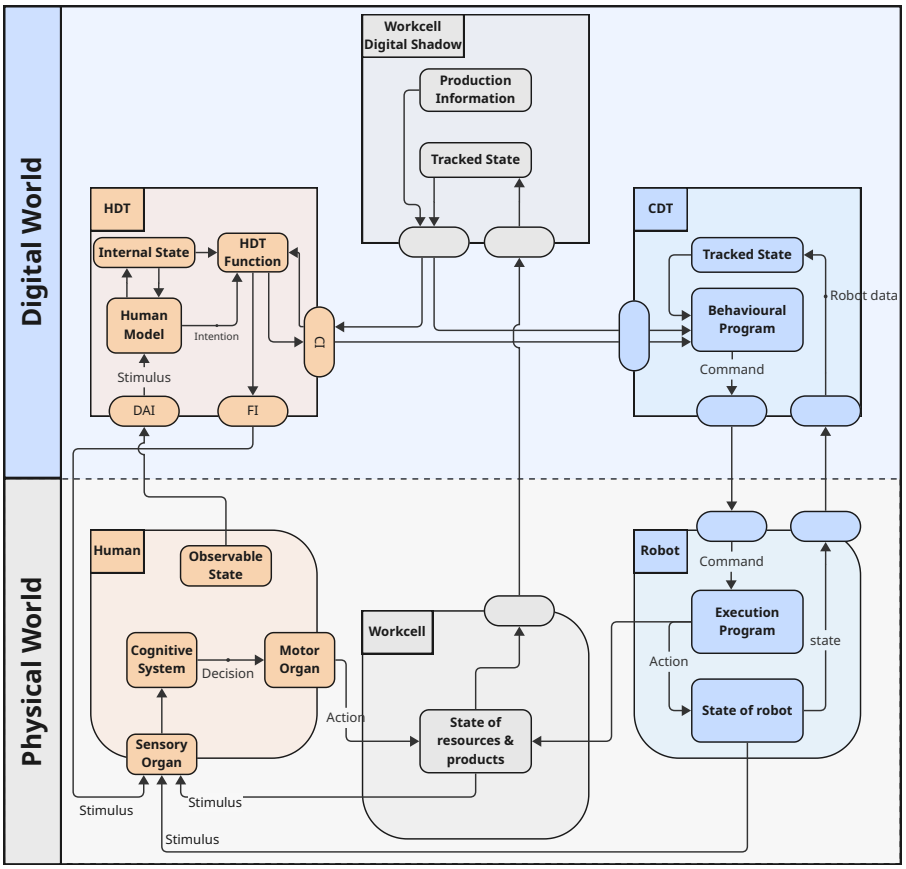


Figure 3.4: Digital Twin based architecture and data flows for HRC scenarios.

them to the other DTs, enabling real-time situational awareness about the concurrent changes in the environment.

In addition, the WDS contains information about the production goal. The production task can be decomposed into subtasks to be coordinated between the collaborating agents. The WDS tracks which subtasks the agents are working on and the overall progress towards the production goal. This positions the WDS as a potential coordinator in collaborative setups. To coordinate subtasks between agent, the requirements to perform a subtask are specified as not only the needed components and tools, but also the abilities required to perform it. To provide a formal grounding to a system's abilities (human or engineered), we refer to the systemic core conceptual model established in our work published in Naudet et al. [2025]. In this framework, a System (human or technical) is defined by its *Properties* and *Functions* while existing within an Environment characterized by mutual influence. An *Ability* is defined as a special property indicating that the system is "capable of" executing a function. When a system acts toward a goal, it performs an *Activity* enabled by its abilities.

Robot and its Digital Twin:

The physical robot contains the required *Execution Program* to execute commands and keep track of its own state. However, the selection of what action to execute is delegated to its CDT. The state of the robot is continuously communicated to the DT where it is tracked and used as input to decide on what action to execute next. The *Behavioural Program* of the CDT is the processing unit corresponding to the human's *Cognitive System*. It receives the required information about the robot & environment state and task requirements to choose the next command. The CDT integrates a cognitive architecture that extends the functionalities of a conventional digital twin beyond state tracking. Specifically, it enables predictive reasoning, adaptive decision-making, and problem-solving [Al Haj Ali et al., 2024a]. This makes it possible not only to choose the robot's next step but also to learn from and adapt to unexpected situations occur, such as missing parts, delays, or task conflicts.

Additional Data Flows:

Beyond the data flows between each entity and its DT, data can also be exchanged between DTs. In general terms this enables DTs to have access to complete contextual information that the physical entity could not have on their own. The WDS shares data about the current state of the environment and activity of other agents in the system with the two other DTs. The HDT and robot's DT can both send and receive information from each other. The access to additional contextual data enables adaptations that purely physical entities could not produce. We highlight three specific instances that are due to the inherent limitations of humans and the blindness to their internal state:

1. **Proactive Robot Adaptation via Intent Sharing:** Current robotic systems lack information about the internal human states, such as intentions. While sensors can detect physical movement, a robot can typically only react once an action has already begun, leading to interrupted workflows and reduced efficiency. By contrast, the proposed architecture allows the HDT to host an emulation model that reflects the human's *Internal State*, including their current goals and task knowledge. When the HDT estimates the next planned action based on the emulation model's output, it can share this unobservable intent with the robot's CDT. This allows the robot's *Behavioural Program* to proactively

avoid conflicts, such as reaching for the same component, before physical movement even occurs, thereby fostering a shared intention necessary for cognitive interoperability.

2. **Selective Contextual Awareness and Information Filtering:** Human agents have a limited point of view and inherent biological constraints, such as limited processing capacity and memory. Within this architecture, the WDS tracks all environmental changes and tool availability in real-time. While a human might miss a task-relevant change outside their immediate field of vision, the HDT receives these updates through its *Communication Interface*. Because the HDT understands the individual's current goal, it can selectively filter this data and provide only goal-relevant information via the HMI. This ensures the human remains aware of relevant environmental shifts without being distracted or overwhelmed by unnecessary data.

3. **Predictive Guidance and Error Prevention:** Traditional industrial systems are governed by pre-defined rules that struggle to handle the uncertainty of human behaviour or unexpected deviations from a task. In such setups, mistakes are often only identified and corrected after they have occurred. The proposed architecture addresses this by linking the HDT to the production goal tracked by the WDS. Since the HDT emulates the human's cognitive processes, it can assess whether an upcoming action aligns with the production sequence or represents a potential error. If the model determines that an action does not advance the production goal, the HDT can provide proactive guidance or warnings. This shifts the collaboration from a reactive safety model to real-time support of the worker's decision making to ensure the successful fulfilment of shared objectives

3.3.2 Use Cases

To illustrate the functioning of the proposed DT-based architecture we describe two specific use cases. The first is about adapting a robot to avoid any *concurrent actions*. The second about providing guidance to a human assembling products.

Use-case 1: Adaptive Robotics - Avoiding conflicting actions

Consider a collaborative assembly task involving one human and one robot. Both have the shared objective of assembling a set of products that requires the execution of a sequence of assembly steps. Each step requires that the agent performing the task has the required skills and access to all necessary components or tools. While the objective is to go towards autonomous and active collaborators, we simplify the present use case by having a freely acting human and adaptive robot.

In this use case, only *independent actions* and *concurrent actions* can occur (see Figure 3.3). Both agents assemble products independently and have all the skills required to do so, but they share the same components and tools. The possible concurrent actions are, both agents take the same component at the same time, or both agents take the same tool at the same time. We refer to these two situations as *resource conflicts*. Applying the DT-based architecture from Figure 3.4 would enable the dynamic replanning of robot actions through the data sharing between HDT and CDT. The roles of each physical and digital entity in this setup are the following:

- **Human Individual** (*Physical Layer*): The human has the goal to advance the production goal by performing goal-oriented actions. The human possesses a set of abilities to perform task-relevant activities. They perceive environment stimuli to form an understanding of the current situation and select the next action to perform.

- **Robot** (*Physical Layer*): A physical agent responsible for the physical execution of assembly steps. It houses an Execution Program that manages low-level command fulfilment and continuously tracks the robot's own physical state and coordinates. It does not make its own decisions but executes commands received from the CDT.
- **The Workcell** (*Physical Layer)*: This is the shared physical environment. It contains all the necessary resources, including tools, components, and the products being assembled, that are used by both the human and the robot. Resource conflicts can occur in this space.
- **Human Digital Twin** (*Digital Layer*): The HDT has the objective of emulating the twinned individual's cognition and behaviour to enrich the available information for robotic adaptations. It hosts an emulation model to model and track the cognitive state including the human's intention.
- **Cognitive Digital Twin** (*Digital Layer*): The CDT hosts a behavioural program to select the next action for the robot. It uses cognitive architectures to enable adaptive decision-making, allowing it to solve problems such as task conflicts dynamically.
- **Workcell Digital Shadow** (*Digital Layer*): Serves as source of information about the workspace. It tracks the state of all resources (tools, components) and product progress, sharing this environmental context with both the HDT and CDT

In this setup, when the human worker intends to reach for a component, the HDT emulates their decision-making process and estimates the next planned action. Through the Communication Interface, the HDT shares this unobservable intent with the robot's CDT before the physical movement occurs. The CDT's Behavioural Program evaluates this intent against its own planned actions and the WDS's resource data, allowing the robot to proactively select an alternative task or path

In absence of the DTs, one may try to solve the problem with additional sensing capabilities for the robot. The physical motion of the human is detectable and can be used to adapt the robot's movement. However, it is only possible to react to an action once physical movement has started. From the movement, the risk of conflict or collision needs to be inferred to subsequently interrupt the robot's current action. Even if all risks of collisions are eliminated, it will necessarily lead to interrupted actions that represent lost time and reduced efficiency of the workcell. In contrast, the DT-based approach aims to avoid selecting any actions that may lead to a conflict before they happen. The key differences between the DT-based approach and the non-DT based approach are shown in Table 3.1.

Use-case 2: Guidance during Assembly Task

Within the AI4C2PS project, we explored a second use case involving a social robot named QT robot and developed by LuxAI[1]. Its objective is to guide a human during an assembly task. An early user study for vocal guidance based on a Retrieval-augmented generation (RAG) LLM system was published in Anastasiou et al. [2025]. The publication discusses plans to deploy an HDT to provide personalised and dynamic guidance by exploiting the modelled knowledge levels and situational awareness of the human. Effectively QT robot is taking the role of an interface between human and HDT.

[1] https://luxai.com/

Table 3.1: Use Case 1: Potential benefits of the DT-based collaborative assembly setup compared to the non-DT based alternative.

	Non-DT Based Setup	**Decentralized DT Setup**
Operational Mode	**Reactive:** The robot is blind to the human's internal state and can only react to detected physical motion via safety sensors.	**Proactive:** The robot adapts to estimated intentions, allowing for seamless coordination before physical conflicts arise.
Workflow Fluency	**Interrupted:** Conflicts (e.g., both agents reaching for one tool) lead to emergency stops or pauses, reducing efficiency.	**Fluid:** The sharing of intentions allows the robot to dynamically replan its actions to maintain production flow.
Trust & Ergonomics	**Stressing:** Sudden reactive stops or conflicting unpredictable robot movements can erode human trust and increase mental workload.	**Supportive:** Predictable, intent-aware robot behaviour could enhance legibility of movement and the worker's perceived safety and trust.

Consider a human assembling a product on their own. They are seated at a workstation that contains all the needed components and tools in a set of drawers. To successfully perform the task, a sequence of assembly steps needs to be executed in the correct order. Each step can only be performed successfully when the requirements are met, which are the availability of the needed tools and components, as well as the human having the required skills. Using the wrong component or wrong tool risks damaging or destroying the partially assembled product.

The instructions are provided on a screen, one step at a time. To move to the next instruction the worker can hit the `next` key on a keyboard. To complete the step the worker needs to consult and understand the instructions, find the needed component and/or tool in a set of drawers, and assemble the component correctly. The HDT is continuously tracking the stimuli sensed by the human. In this case there are two types of stimuli which can both be detected relatively easily. The first is the visual information shown on screen when the worker consults the instructions. The second type is the content of a drawer when it is opened. Throughout the task the emulation model processes the stimuli to update its internal state and estimate responses. The internal state includes the memory of the drawer contents that is progressively reinforced and the knowledge about the instructions. Thanks to the emulation model's internal state and output the HDT can send personalised guidance to the worker when needed. We foresee two helpful types of guidance this setup enables:

1. The HDT can interrupt when the estimated next action risks damaging the product.
2. The HDT can remind the location of task-relevant components and tools or details about instructions when the memory has decayed.

In absence of the HDT, there is still the possibility to provide guidance. For the first case an action can be interrupted, only when the action has already started and a drawer opening has been detected. Interrupting in this reactive way may be fast enough to avoid damage, but constant interruptions while the worker is searching for a component may be perceived as disturbing and unpleasant. For the second case an assistant could provide guidance every time

the next step in the instructions is consulted. This may lead to a dependence on the system, as the worker is never incentivized to learn the setup around them in full. In contrast the HDT may reinforce memories selectively to promote learning task-relevant information in full. Table 3.2 contrasts the potential benefits of the HDT based setup with a traditional guidance system.

Table 3.2: Use Case 2: Potential advantages of an HDT-based guidance system during an assembly task compared to a non-HDT based alternative.

	Non-HDT Based Setup	**HDT Setup**
Error Handling	**Reactive:** Mistakes are identified after they occur (e.g., detecting an open wrong drawer), often resulting in damaged parts.	**Proactive:** Errors are anticipated by emulating the worker's intent, allowing for intervention before the physical act.
Information Flow	**Generic/Constant:** Instructions are often provided at every step, which can cause dependency or distraction.	**Selective:** Guidance is filtered based on the individual's internal goals and predicted knowledge levels, providing help only when needed.

Scalability of the Architecture

The architecture proposed in this chapter adopts a decentralised approach where each entity has their own dedicated DT (see Figure 3.4). While the discussed use cases are limited to two agents, the architecture is designed with scalability in mind. The decentralised approach enables dynamic expansion of the multi-DT system as more agents enter the shared workspace. By creating persistent, personalized HDTs that accompany individuals across contexts, the architecture moves away from rigid, location-bound representations toward a flexible digital environment where multiple HDTs, CDTs, and Workcell Digital Shadow (WDS) interact. In principle, an arbitrary number of DTs can connect and communicate through their respective *Communication Interfaces* which enables them to negotiate and make decisions on behalf of their twinned entities. The complexity induced by this multi-agent setup leads to specific challenges. As the number of collaborating agents grows, so does the risk of entering into conflict and erode the quality of collaboration. Indeed, coordination becomes a central concern which may be addressed with approaches from Multi-Agent Systems research [Wooldridge, 2009].

The DTs can be considered as agents that have the objective to support their twinned entity by acting on their behalf in digital space where decisions such as task allocation are made. For task allocation *market-based mechanisms* [Wellman, 1993] could be used where each DT bids on behalf of their twinned entity based on estimated performances (execution time, resource usage, confidence scores). *Consensus mechanisms* [Olfati-Saber and Murray, 2004] are of particular interest in situations where agents have different perspectives and act on limited information. A typical example are voting mechanisms, where each agent votes on the next steps to take in a collaborative plan. In our case the DTs handle the voting and joint decision making, while the twinned entities execute the plan. However, they can also be used to resolve conflicts in understanding of the state of the environment, which makes them particularly interesting to achieve cognitive interoperability.

As humans aren't automatons, they may disagree with decisions made by the DTs and refuse to execute them. Therefore, an additional procedure may be required to take into account human refusal and start subsequent negotiation rounds that respect this new constraint. This framework

extends to any collaboration between humans and machines. While our discussion is anchored in collaborative robotics, it is extendable to CPS in general. Within the industrial context, it may include Autonomous Intelligent Vehicles (AIV), control systems, and automated workstations.

3.3.3 Challenges

The DT-based approach creates opportunities to improve HRC by exploiting information about the human's cognitive state and processes. A central pillar is the exploitation of an emulation model for human cognition that is deployed in an HDT. By processing the same stimuli as the twinned human, it should keep an aligned representation of the human's cognitive state. Accessing this information about, among others, the current goal and task-relevant knowledge, is the key enabler for cognitive interoperability and improved collaboration. To move towards implementation and demonstrate the feasibility of the approach creation of said emulation model comes with a set of challenges.

- **Modelling Cognition:** Emulating cognition and estimating the actions based on stimuli requires a model of cognition that includes all cognitive processes from perception to action selection. The inherent complexity of human cognition and partial understanding of how the brain works limits the fidelity of models. Additionally inter-personal differences may impact behaviour leading to the need of personalised emulation models for HDTs.
- **Sensing Stimuli:** To emulate the cognitive process of a human, the physical stimuli need to be detected. The detection difficulty depends on the modality. For aural stimuli a microphone in the proximity of the twinned human may be sufficient. Any noise that is picked up by it, may also be audible to the person. For visual stimuli the difficulty is significantly higher as many objects within a human's field of view may compete for attention. Therefore, information about the gaze is required. Eye-trackers provide information about visual attention by pinpointing what objects the human is exactly looking at.
- **Lack of observability of Internal States:** Unlike physical movement, which can be sensed through external sensors, the internal cognitive states are mostly non-observable. This hampers the ability to verify whether a deployed emulation model is correctly representing the internal cognitive state of its twin. Direct detection methods, such as EEGs or fMRIs, are often invasive and impractical for industrial work environments. This blindness to the internal state is a critical barrier to cognitive interoperability.
- **Real-time Requirement:** For use cases that require quick adaptations of the robot, the deployed model needs to run in real-time. Consider the presented case for robotic adaptation where the human will receive a stimulus and decide within less than a second what action to take. The entire pipeline from detecting a stimulus to adapting the robot behaviour would need to run faster than the human's decision to act. This includes predicting the next action and running the behavioural program of the robot's DT.

The challenges identified above represent critical roadblocks to achieving cognitive interoperability and enhancing HRC. While the creation of emulation models for human cognition is the cornerstone of the proposed HDT-based approach, it also remains the primary source of modelling difficulty. In the next sections, this thesis addresses this central obstacle by formalizing the essential properties and internal dynamics required for cognitive emulation. Ending with an exploration of existing cognitive modelling approaches to identify the most promising architecture for emulating human cognition in collaborative industrial contexts.

3.4 Emulation Model for Human Cognition

Cognitive interoperability requires the continuous sharing of internal cognitive states to avoid misunderstandings and breaking down of collaboration. Workers may self-report their intentions, reasoning, and actions continuously to grant collaborating systems insights into their internal processes, but this represents a disruptive and highly unnatural way of interaction. Instead, HDTs can act as an intermediary that model the worker's cognitive process and communicate intentions and understandings with collaborating systems. The HDT's structure is designed to gather data continuously, model the cognitive processes and communicate relevant decisions, or states, with external system. Modelling these cognitive processes is a key challenge. Behavioural differences between individuals are important, making general population-wide models unlikely to succeed. An HDT is unique for the twinned individual, being entirely specialised on not only receiving data from that individual but also hosting personalised models.

The heart of an HDT is its emulation model that produces the same *input-output* relation than the twinned individual it represents. Within the context of human cognition, the input is a stimulus coming from the environment and the output is an action. A wide variety of modelling approaches can be used to map incoming stimuli onto resulting actions. This includes data-driven approaches from the field of Machine Learning and Artificial Intelligence that often lack explainability and interpretability [Linardatos et al., 2020]. However, cognitive interoperability requires information about intentions and mental models making black box models that map stimuli to motor actions insufficient. Hence, a stateful model is required to enable cognitive interoperability. Later in section 3.4.2, we explore approaches from cognitive modelling to identify a suitable model for the emulation of human cognition.

3.4.1 Emulation Model Formalisation

Figure 3.5 shows the relation between HDT, twinned human and the hosted emulation model alongside the data flows. It follows the structure of the meta-model from section 2.3.3, but some elements are omitted for clarity. A stimulus is sensed by a human who processes it to take a decision and execute a motor action. The same stimulus is sent to the HDT where it is processed by the emulation model. The emulation model executes a sequence of internal processes representing human cognition until the next action is selected. During the processing the internal state is updated. If additional stimuli are created by external sources, they may impact the chain internal processes and the selected motor action. The estimated action can serve as input for HDT functions to generate feedback for the twinned human (see section 2.3.7 and section 2.3.3 for details). The motor action selected by the twinned human changes the state of the environment, potentially leading to the generation of new stimuli. A successful emulation model needs to estimate the same motor actions as the twinned human and keep its internal state aligns with the human cognitive state.

We mathematically formalise the model and its relation with the environment. For this formalisation we consider a *single agent* (the modelled human) *interacting with the environment*, where all changes in the environment are caused by the actions of the agent. Hence, there are no actions from external systems or spontaneous processes that may interrupt the internal processes of the model. We define the *Task Environment* as a tuple $(\Sigma, O, A, T, \mathbb{O})$ where:

- Σ is the set of environment states
- O is the set of observable stimuli

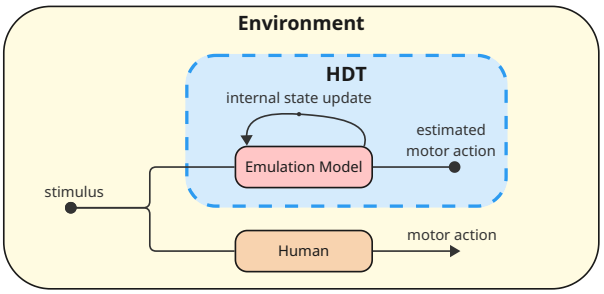


Figure 3.5: Basic input-output relation of an emulation model for human cognition in relation to the physical environment.

- A is the set of physical actions
- $T : \Sigma \times A \to \Sigma$ is the state transition function for the environment
- $\mathbb{O} : \Sigma \to O$ is the observation function

The observation function $\mathbb{O}$ takes an environment state as input and outputs the observable stimuli. This is necessary as parts of the environment state can be hidden from the agent's viewpoint. The transition function T resolves the action of the agent and updates the environment state. We assume that the transition function T and observation function $\mathbb{O}$ are deterministic, producing a single output. This means that once an action is selected its execution always leads to the same change in environment state. The same applied to the observable stimuli that result from a given environment state. Note that the stimuli only represent the information that can be observed by the human, it does not give full information on whether the human pays attention to the stimuli or how it is processed.

Let $\mathcal{M}_\Theta$ be the emulation model defined by a tuple (S, Θ) where S is the set of internal states and Θ is the set of parameters driving the model's internal processes. The emulation model interacts with the task environment by processing observable stimuli $o \in O$ and selected the actions to execute $a \in A$. During the processing the internal state $s \in S$ is updated as follows:

$$\mathcal{M}_\Theta : O \times S \to A \times S \tag{3.1}$$

Note that the above equation allows for the incoming observed stimuli to be empty $o_t = \emptyset$, as well as the selected action $a_t = \emptyset$. This represents a case where the model is internally processing information and reasoning about what action to take next. The impact of inter-personal differences can be captured by personalising the model through its parameters Θ. Indeed, for HDTs the model should be personalised and unique to the twinned individual. Figure 3.6 provides an overview of the interaction loop between environment and model.

The model can be used for various purposes, including simulation and emulation. It is important to make a clear distinction between those two functions. Simulations take a hypothetical initial state and execute the model to explore and quantify the response. This makes simulation a function that is disconnected from the physical entity but serves as important tool to explore model behaviour and limitations. By contrast emulation is tightly interconnected with the physical entity. It produces the same input-output behaviour as the modelled entity. When deployed in a DT the emulation model is synchronised with the real-world behaviour of the twinned system.

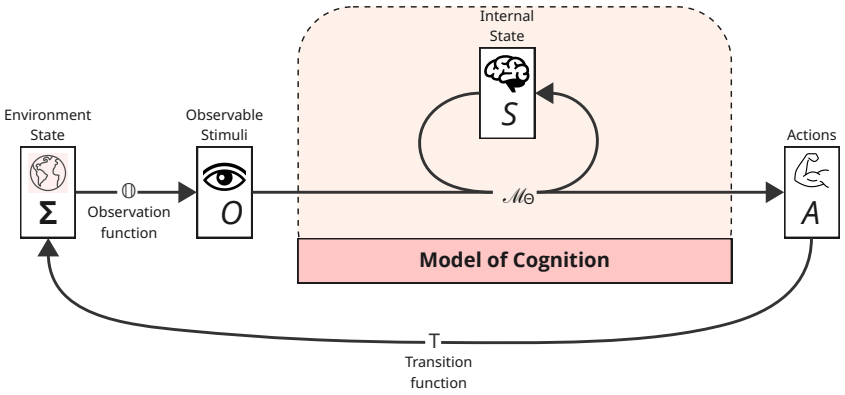


Figure 3.6: Relation between the environment, model and various processes.

Algorithm 1 details how emulation can be implemented based on the formalisation. A key feature is the synchronisation with the observed actions taken in the real-world, hence an observed action trace $a = (a_0, ..., a_N)$ is required. From an initial environment state σ_0 and internal state s_0 the observable stimuli o_k is computed from the environment state σ_k. Before executing the model, a copy of the internal state is saved in a separate variable s_i. The model in executed until it selects the next action $\hat{a}_k$. If the selected action $\hat{a}_k$ is different from the observed action a_k the synchronisation function $g(s_i, a_k)$ is applied. Finally, the environment state is updated based on the observed action a_k. This procedure is continued until the sequence of observed actions ends. In real-time settings the observed actions would be received continuously instead of being a finite sequence.

Algorithm 1 Emulation

Require:
Initial environment state $\sigma_0 \in \sigma$
Initial model state $s_0 \in S$.
Sequence of observed actions $a = (a_0, ..., a_N)$

Ensure:
$\sigma_k \leftarrow \sigma_0$
$s_k \leftarrow s_0$
for a_k in a **do**
  $o_k = \mathbb{O}(\sigma_k)$
  $s_i \leftarrow s_k$ ▷ Copy of initial state for synchronisation
  while $\hat{a}_k = \emptyset$ **do** ▷ Run model until action is selected
    $(s_{k+1}, \hat{a}_k) = \mathcal{M}_\Theta(o_k, s_k)$
    $s_k \leftarrow s_{k+1}$
  end while
  if $\hat{a}_k \neq a_k$ **then** ▷ Only synchronize if selected action was wrong
    $s_{k+1} \leftarrow g(s_i, a_k)$ ▷ Infer internal state from observed action
  end if
  $\sigma_{k+1} = T(\sigma_k, a_k)$
  $s_k \leftarrow s_{k+1}$
  $\sigma_k \leftarrow \sigma_{k+1}$
end for

Inferring the internal state from an observed action is a critical step for synchronisation and emulation. However, there may be many sequences of internal processes that lead to the same observed action, but different internal states.

Algorithm 2 details how the formalisation can be used to simulate the execution of a task. To run a simulation the initial state of the environment σ_0 and model s_0 are needed. Additionally, the stop condition needs to be defined. The simulation stops when the task is considered finished by reaching the desired target state of the environment σ_{target}. For instance, for an assembly task the desired environment state contains the fully assembled product. The core loop of the simulation is (i) computing the observable stimuli o_k from the environment state σ_k, (ii) computing the selected action a_k and updated internal state s_{k+1}, and (iii) resolving the action a_k to compute the updated environment state σ_{k+1}. The updated environment state σ_{k+1} and internal state s_{k+1} forms the starting point for the next iteration. The output of a simulation run is an action trace $\hat{a} = (\hat{a}_0, ..., \hat{a}_N)$ that describes the simulated behaviour during task execution. This trace can be used to evaluate the model by comparing it to observed human action traces. In particular, from the trace the performances at fulfilling the task can be extracted and compared to real human performances. This provides valuable indication on whether the model is able to reproduce human flaws, instead of executing the task optimally.

Algorithm 2 Simulation

Require:
Initial environment state $\sigma_0 \in \sigma$
Initial model state $s_0 \in S$.
Target environment state σ_{target}
Ensure:
$\sigma_k \leftarrow \sigma_0$
$s_k \leftarrow s_0$
while $\sigma_k \neq \sigma_{target}$ **do**
$o_k = \mathbb{O}(\sigma_k)$
$(s_{k+1}, \hat{a}_k) = \mathcal{M}_\Theta(o_k, s_k)$
$\sigma_{k+1} = T(\sigma_k, \hat{a}_k)$
$\sigma_k \leftarrow \sigma_{k+1}$
$s_k \leftarrow s_{k+1}$
end while

With this mathematical formalisation, we clarified the interaction between model and environment. This provides a useful framework when moving towards implementation. In the next section, we explore the existing modelling approaches to select the best model for emulation of human cognition.

3.4.2 Cognitive Modelling

The keyword cognition is understood differently within various contexts. To clarify our use of the word, we refer to a definition from cognitive psychology: "*Cognition is all the processes by which the sensory input is transformed, reduced, elaborated, stored, recovered, and used*" [Neisser, 1967]. This definition includes a large variety of cognitive functions, including the lower level functions like perception or memory and higher cognitive functions including meta-cognition.

We've published a hierarchical model of the different components of the cognitive system in Naudet et al. [2025].

The motivation behind modelling cognition can be very different. There is the scientific quest of understanding how the human cognition works and gives rise to our abilities to reason, learn and live. But there is also the desire to create artificially intelligent systems with cognitive abilities inspired by the understanding of our own mind. As a consequence, both the field of cognitive science and artificial intelligence research intersect. Early ambitions to create Artificial General Intelligence came from cognitive science where the hypothesis was made that the human mind is an information processor and can be replicated as a computerized model [Newell and Simon, 1975; Fodor, 1975].

Bayesian and probabilistic Models: Bayesian models assume that significant parts of cognition can be formalized as problems of induction i.e. making conclusions about the world and its mechanisms based on limited observations and use the gained knowledge to make predictions. Bayesian interference is proposed as the probabilistic framework to model inductive reasoning [L Griffiths et al., 2008]. It assumes that the empirical observations are the result of underlying mechanisms that can be inferred. It requires prior knowledge about the world that is refined as observational data is encountered. The approach is used to model various cognitive processes for example; inferring object properties from natural images [Yuille and Kersten, 2006]; inferring rules of grammar from phrase structure [Chater and Manning, 2006]; inferring causal relationships [Steyvers et al., 2003]. Bayesian models focus on problems of induction addressing the question of how humans build models of the world from sparse and noisy sensory data. The approach applies to various cognitive functions but does not create a single unified architecture of the mind.

Deep Learning & Connectionism: Inspired by the structure of neurons in biological brains, the concept of using computerized artificial neural networks (ANN) has lived through major successes since the introduction or the Perceptron in 1958 [Rosenblatt, 1958]. ANNs are composed of a large number of tunable parameters referred to as weights that are adapted based on example data. Since the introduction of gradient-based backpropagation [Rumelhart et al., 1986] an ever larger class of tasks could be learned by ANNs. The approach has also been used in the field of cognition. However, Deep Learning models are widely considered black box models and lack explainability, i.e., the parts of the model that drive a decision cannot easily be diagnosed [Ras et al., 2022]. The study of human mental processes using ANNs, or connected parallel processing units, is known as *Connectionism* and assumes that cognition emerges from the interaction of these simple units. ANNs are also at the center of the *neuroconnectionist* research programme, which attempts to model behavioural and neural data [Doerig et al., 2023]. *Neuroconnectionism* sets itself the goal to bridge the gap between low level sensory data and high-level goal-directed behaviour. It recognizes the enormous complexity and limited explainability of ANNs but argues that this does not represent a fundamental roadblock. They argue that ANNs should be considered "transparent boxes" because all weights are accessible and can their effect can be studied by for example lesion studies. Sequence based models have been used to predict decisions, an example is the LSTM architecture for reward-based tasks [Fintz et al., 2022]. Finally, Large Language Models (LLMs) have shown an unexpected capability to not only process language but sequential data in general. They have been used to create small artificial agent that exhibit human-like behaviour, like forgetting to show up to a social gathering [Park et al., 2023]. LLMs have been used for multi-model data including vision, and their outputs used for robotic control.

Driess et al. [2023] have demonstrated that this approach can yield promising results for multiple types of tasks with a single model. The ambitions of these models is more to create intelligent cognitive agents than to exactly reproduce the behaviour of humans. It is important to note that LLMs are fundamentally next-token predictors without an internal structure that allows for reasoning in the human sense. Thus, claims of reasoning capabilities should be taken with a pinch of salt.

Cognitive Architectures are unified and cohesive frameworks that model multiple cognitive functions within a single system. They are based on the theoretical understanding about the human mind and aim to recreate its functioning in a computerized system. There is a large number of cognitive architectures that differ from each other in their theoretical assumptions, computational mechanisms and biological plausibility. Three main categories are ***(i)*** Symbolic architectures rely on rule-based systems to model cognitive processes. It follows the assumption that the human mind is in essence manipulating symbols and can be recreated in silica if all rules are known and encoded; ***(ii)*** Emergent or Connectionist architectures use distributed sub-symbolic processing, inspired by neural networks and biological systems. Their efforts are based on the assumption that complex cognitive processes emerge from the interaction between simpler units; and ***(iii)*** Hybrid architectures combine symbolic rules with the distributed sub-symbolic processing. Cognitive architectures have not only been used to model human behaviour but also as create cognitive artificial systems. They have for example been used as models of robotic control to enable more flexible behaviour.

For an HDT specifically we've argued in Naudet et al. [2024], that the emulation function should be implemented using a cognitive architecture to accurately replicate the human mind's processes. This is because they are inherently designed to build a unified model of cognition by describing the fixed mechanisms and underlying structures that support human performance across multiple tasks and modalities [Newell, 1990; Laird et al., 2017]. This provides a necessary distinction between mere "human-like behaviour" and true "human-like cognition," where the HDT follows the same cognitive cycle (perception, attention, memory, and reasoning) as its physical twin. Ultimately, this structural and process-level alignment is what may enable cognitive interoperability, as it allows artificial agents to align their perceptions and shared intentions with humans through models that are inherently interpretable and grounded in human-like logic. In the next section, we present the key elements that constitute a cognitive architecture and select the most fitting one for our subsequent experiments of the final chapter.

3.4.3 Cognitive Architectures

"Even if the mind has parts, modules, components, or whatever, they all mesh together to produce behavior. ... If a theory covers only one part or component, it flirts with trouble from the start." [Newell, 1990, p.17]

In the 1970s the field of psychology reached a level of maturity that gave rise to successful task-specific models for cognition. This led to a landscape of highly specialized models and methodologies, but no integrated model of cognition. As a response, Allen Newell proposed the development of Unified Theories of Cognition. He describes these theories as *"[...] Theories that gain their power by positing a single system of mechanisms that operate together to produce the full range of human cognition"* [Newell, 1990]. The efforts towards unified theories of cognition lead to the development of cognitive architectures. They are specific structural implementations designed to create executable models and test its predictions. Any cognitive architecture implic-

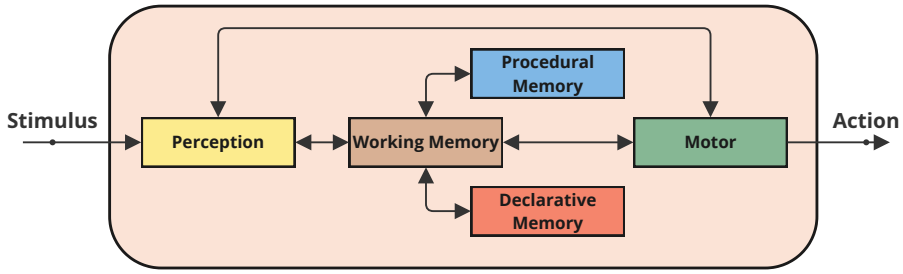


Figure 3.7: Common Model of Cognition from Laird et al. [2017].

itly includes assumptions about the fixed structures of the human mind. Over multiple decades a vast variety of cognitive architectures were developed with different starting assumptions and implementation approaches [Kotseruba and Tsotsos, 2020]. In 2017, Laird et al. [2017] analyse multiple well-established cognitive architectures to synthesise the core components and assumptions. It is important to note that the standard model seeks to model human like minds which includes their inherent limitations. The purpose of the model is not optimality but bounded rationality. Bounded rationality describes the inherent biological constrains of human cognition, including incomplete knowledge, limited processing capacity and limited memory.

In Figure 3.7, a high-level view of the core components is shown including; *Perception* and *Motor* modules; *Working Memory*; and two types of long-term memory, namely *Procedural Memory* and *Declarative Memory*. Each module can be further decomposed, for example the *Perception* module is often composed of one sub-module per modality, such as vision or sound. *Perception* and *Motor* module have a direct connection, while all other modules rely on the *Working Memory* as central communication bridge. Communication runs through buffers, with each module only having access to their associated buffer. The only exception is *Procedural Memory*, which can access all buffers.

The central concept of the standard model is the *cognitive cycle*. *Procedural Memory* selects a single deliberate act per cycle based on the buffer content and induce changes in them. A deliberate act includes internal processes such as a step of reasoning, retrieval from long-term memory, as well as the initiation of motor actions. Only a single deliberate act can be selected per cycle, representing a bottleneck for human performance. The duration of a cognitive cycle is approximately 50ms. While *Procedural Memory* imposes a sequential selection of deliberate acts, while all other modules execute their own processes in parallel. Complex behaviour arises from the sequential execution of cognitive cycles without any additional module for global optimization. *Long-Term Memory*, both procedural and declarative, contains symbol structures and associated quantitative metadata that influences its processing. For examples the retrieval of declarative memory can be associated with an activation that impacts retrieval probability and speed. Please note, that *Procedural Memory* is acting upon symbols, making the core of the architecture symbolic in nature. Nonetheless, sub-symbolic processes modulate the selection of the deliberate act. All forms of learning occur incrementally and online. This includes the acquisition and reinforcing of long-term *declarative memory*. For *procedural memory*, rules are reinforced when they lead to increased performance. Additionally, multiple rules can be merged into a single one, representing behavioural automatization.

The standard model provides a set of assumptions to create biologically plausible cognitive architectures. It is composed of multiple modules that are connected through the working memory. The state of the working memory and modules represents the internal state of the modelled

individual. This corresponds to the internal state s we defined for emulation models in Section 3.4. Additionally, the content of the working memory buffers is symbolic in nature, enabling interpretability of the state. In our formalization the emulation model $\mathcal{M}_\Theta$ updates the internal state and/or selects an action. Conceptually this maps onto the cognitive cycle of the standard model and the selection of a deliberate act.

Cognitive Architecture Selection

Cognitive architectures are well aligned with our needs for emulating human cognition; they model the cognitive functions needed to interact with an environment in a single computerized system. Since their inception, the number of cognitive architectures has grown enormously. This ranges from small scale conceptual work to multi decade efforts to build up fully integrated architectures. For our experiments we need an architecture that can reproduce human behaviour, including the limitations of human cognition, with the highest degree of plausibility and accuracy. We translate this into a set of requirements for the architecture selection:

- **R1:** The architecture is biologically plausible.
- **R2:** The architecture can process stimuli and select motor actions.
- **R3:** The architecture's internal state is accessible and interpretable.
- **R4:** A library exists to create models and integrate them into larger systems.

Kotseruba and Tsotsos [2020] surveyed a set of 84 architectures to describe their core abilities without strictly constraining their work to architectures that follow the standard model of the mind [Laird et al., 2017]. Architectures are categorized based on how information is represented and processed. Three major paradigms are distinguished; ***(i)*** symbolic architectures that represent information as symbols and apply if-then rules on them; ***(ii)*** emergent approaches, also known as connectionist, are massively parallel models, such as neural networks; ***(iii)*** hybrid architectures seek to combine both approaches. Each paradigm comes with its own advantages and drawbacks. For our selection process the following are important to consider. Purely symbolic architectures lack the flexibility in their perceptual processing to deal with dynamic environments. On the other hand, emergent approaches learn from data but lack the interpretability we seek.

For a more detailed analysis, we select the five of the most widely used architectures based on the review by Kotseruba and Tsotsos [2020]. We only consider architectures with a track record of continuous development that have not been abandoned, with published application cases in either Human-Computer Interaction or Robotics. Namely, we take a closer look at Leabra [O'Reilly et al., 2012], CLARION [Sun, 2006], ACT-R [Anderson et al., 2004], Soar [Laird, 2019] and SPA/SPAUN [Stewart et al., 2012] before selecting an architecture for implementation.

Leabra is an emergent architecture that implements biologically realistic neural processing mechanisms. Its focus is on visual perception by replicating the structure of the ventral stream [O'Reilly et al., 2012]. The ventral stream is the part of the brain that processes the visual signal and identifies what objects are in the field of view. Motor control is limited on eye-movement to reproduce realistic saccades and shifting of visual attention. The theoretical framework of the architecture is holistic and comprehensive. It is structured across four levels (micro, meso, macro, meta) and contains all required components to create any cognitive behaviour. While implementations focus on visual tasks, the architecture has also been used to implement a sense-making

and reasoning model within the scope of the ICArUS project. Leabra is a purely emergent architecture which makes interpretability challenging. A Go implementation with Python interface is publicly available to develop models.

Clarion has the goal of reproducing human cognition and achieve cognitive realism especially for social behaviour. Its design draws from psychological experiments on social cognition, but is less focused on physiological validation (e.g., mapping functions to distinct brain regions). It distinguishes between explicit and implicit cognition [Sun, 2006], where explicit knowledge is represented symbolically (top level) and implicit knowledge is represented in a distributed sub-symbolic fashion (bottom level). The explicit layer is contributing to the interpretability of the model state. The architecture is further divided into four sub-systems; Action Centered Subsystem to control actions and internal mental operations; Non-Action-Centered Subsystem to maintain and retrieve general knowledge; Motivational Subsystem to model goals and drives towards action; and the Metacognitive subsystem to monitor and improve the performance of all other subsystems. The architecture can interact with the environment through its Motivational subsystem for perception, and Action-Centered Subsystem for action selection. Both perception and motor execution are focused on high-level decision making and not on high fidelity motor control.

ACT-R is a hybrid cognitive architecture that combines symbolic production rules and information chunks with sub-symbolic processes [Anderson et al., 2004]. Thanks to the symbolic core its internal state remains interpretable, while the sub-symbolic processes make it more flexible and biologically plausible than purely symbolic alternatives. It fully follows the assumptions of the standard model of cognition [Laird et al., 2017]. The model can interact with the environment through its perception modules to encode stimuli and motor module to execute actions. ACT-R has been extensively validated on experimental data ranging from behavioural observations to fMRI and EEG response data. It has a dedicated perceptual module and a motor module, but is limited to tasks involving screens, a keyboard and a mouse. A software is available to run and analyse models, which are written in common LISP. An alternative partial Python implementation is available under the name `pyactr`.

Soar is a hybrid architecture with a focus on creating intelligent agents rather than modelling human cognition. It has strong routes in the artificial intelligence community and has been used, among others, for robot control. It has a central working memory that maintains symbolic representations of the sensed data, current goals, and interpretation of the situation. The processing follows two separate phases; first, an elaboration phase where memories are retrieved and potential actions are evaluated; second, a decision procedure that examines the entire retrieved context to select the next action or state change. Soar can perceive the environment with a dedicated Spatial Visual System which relies on a scene graph representation to extract high level symbolic structures [Laird et al., 2012]. To process raw image data an external software is needed to create the scene graph representation. For motor control, Soar has a controller module to transform high level symbolic actions into continuous motor control.

SPA/SPAUN is an emergent architecture that attempts to mimic biological neurons with spiking neural networks, which is a type of artificial neural network. This direct inspiration from low level brain processes leads to a high degree of biological plausibility. However, this leads to a highly distributed and hard to interpret internal state. It has been used to create a model of

the brain called SPAUN, that consists of 2.5 million neurons and can perform 8 different tasks [Eliasmith et al., 2012]. It is able to perceive the environment by compressing sensory input into neural representations referred to as semantic pointers. The architecture manipulates the semantic pointers and selects actions when predetermined conditions are met to execute a rule. High level actions are decompressed into commands into executable joint-level motor commands. The cognitive architecture is implemented with the Nengo software library which is specialized on neural networks.

Table 3.3: Comparison of shortlisted cognitive architectures as they fulfil emulation requirements. (R1: biological plausibility, R2: environment interaction, R3: interpretable, R4: library for integration)

Architecture	**R1**	**R2**	**R3**	**R4**
ACT-R	High	Moderate	Yes	Yes
Clarion	Moderate	Low	Yes	Yes
Leabra	High	Low	No	Yes
Soar	Low	Moderate	Yes	Yes
SPA/SPAUN	High	Moderate	No	Yes

Table 3.3 summarizes how well each architecture meets the requirements we set at the start of the section. First and foremost, the architecture needs to be able to recreate human behaviour accurately and be biological plausible. This criterion allows us to exclude Soar and Clarion from the selection. While Soar is following the structure of the human mind its goal is to create artificial agents with human inspired cognitive functions, rather than reproducing human behaviour accurately. Clarion is interested in social cognition and is psychologically plausible but does not concern itself with the physiology of the human brain. The three remaining architectures have the explicitly goal to reproduce the human mind as accurately as possible but adopt different approaches. Both Leabra and SPA/SPAUN approach the problem from a bottom-up point of view. They seek to reproduce the brain's dynamics by implementing biologically inspired neural networks. The drawback is that their approach is purely emergent and limits interpretability of high-level cognitive states. Finally, ACT-R has high biological plausibility as its modules have been mapped to the brain and validate by brain imaging. It is a hybrid architecture with a solid symbolic core that remains interpretable. The interaction is limited to screen and keyboard-based tasks, but perceptual information can be directly encoded into the architecture, or custom modules can be developed. With this analysis we identify ACT-R as the best match for our experimental needs.

3.5 Conclusions

This chapter focused on the second research question of the thesis (RQ2); *How can HDTs be used to track and emulate human cognition and behaviour? In particular, how can this improve collaboration between humans and Cyber-Physical Systems?*

First, the challenges of Human-Robot Collaboration were presented alongside the context of Cognitive Interoperability. We identified a lack of awareness about the human state, especially their cognitive state, as a particularly important obstacle. To achieve human-awareness, we propose a DT-based architecture for HRC setups with one DT per entity (*Contribution C3*). Here our entities are the human worker, the robot, and the workcell itself. Through the interlinking of

the three DTs new adaptive behaviour is enabled. The exact data flows and exploitation of the HDT is described and illustrated with two use cases. First, a HRC scenario where the objective is to minimize conflicts between human and robot. Second, an assembly case where a human receives personalised guidance based on their specific needs. Finally, we discuss the challenges of implementing the DT-based architecture.

From the challenges, the creation of an emulation model for human cognition sticks out as the most important roadblock. This motivates our efforts to mathematically formalize emulation model for human cognition and its interaction with the environment (*Contribution C4*). We provide two algorithms to use the model for either emulation or simulation. As final part of this chapter, the landscape cognitive modelling approaches is explored to identify the best candidate model for our subsequent implementation. The selected model is ACT-R, as it is the most mature model to simulate human behaviour and has been extensively validated on decades of empirical observations. The next Chapter presents an experimental campaign to assess the suitability of ACT-R to take the role of an emulation model during an assembly task.

Chapter 4

Experiments

4.1 Introduction

In earlier chapters, we established the fundamentals of Human Digital Twins (HDT) and their potential to improve human-CPS collaboration (HCC). We identified the lack of information on the human cognitive state and behaviour as a key challenge for Human Robot Collaboration (HRC) and Cognitive Interoperability. HDTs are well positioned to respond to this challenge by emulating behaviour of the twinned human individual. The heart of the HDT is its emulation model. It models the behaviour and cognitive state of the twinned individual to estimate their next action. This chapter is dedicated to **RQ3** of the thesis:

"Can cognitive architectures model human cognition and behaviour sufficiently closely to be used as emulation model for Human Digital Twins?"

In section 3.4.2 and section 3.4.3 we explored cognitive modelling approaches and identified ACT-R as promising cognitive architecture. ACT-R successfully reproduces the performance and behaviour of humans during a variety of tasks including cognitive test tasks [Anderson and Reder, 1999; Daily et al., 2001], problem solving and decision making [Marewski and Mehlhorn, 2011], skill acquisition and learning [Lebiere, 1999; Taatgen, 2002]. Typically, ACT-R is used to simulate human performance, such as speed and success rates, and evaluated against aggregate data from samples of participants performing the same task [Ritter et al., 2019]. It is essential to clearly distinguish between *simulation* and *emulation* for HDTs. Emulation reproduces the same responses as the modelled system in real-time, while simulation predicts future states or can be used explore hypothetical situations (see section 2.3.7 for more details).

For the ACT-R two questions remain open: ***(i)*** Can ACT-R *simulate* human behaviour during industrial assembly tasks? ***(ii)*** Can ACT-R accurately *emulate* the behaviour of an individual at a response-to-response granularity. This chapter presents two experiments to answer both questions:

1. Section 4.4 presents an ACT-R simulation model for an assembly task. It investigates whether ACT-R can simulate plausible human-like learning and forgetting behaviour during the assembly. We analyse the resulting performance and identify its limitations to propose a set of approaches to improve the model. With this experiment we seek to first establish ACT-R's transferability to the industrial domain before addressing the challenge of using it for emulation purposes.

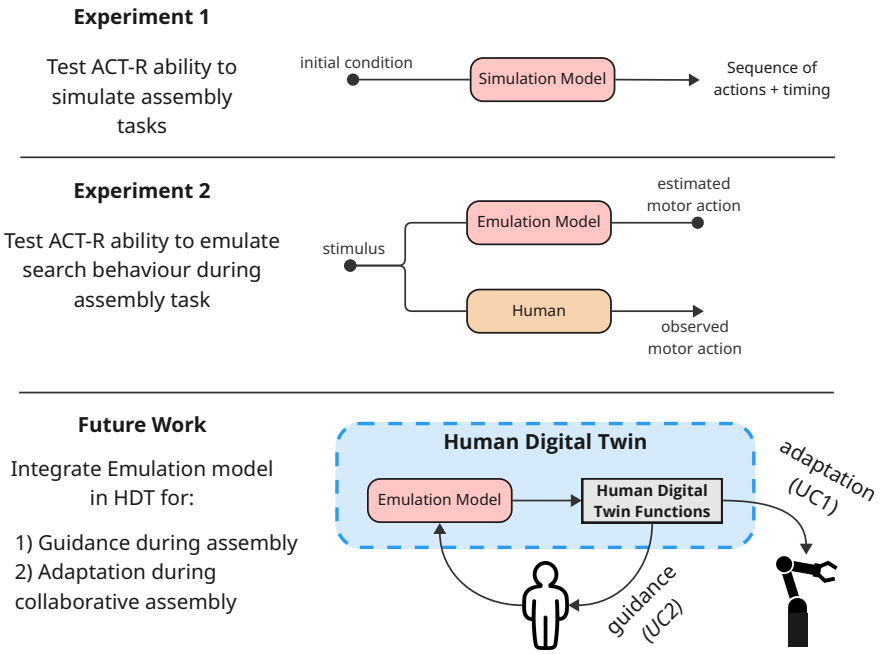


Figure 4.1: Experiment overview and connection to HDT development.

2. Section 4.5 presents an ACT-R emulation model to reproduce the success or failure of individual recall attempts that drive the decision making during the assembly task of section 4.4. To emulate human behaviour, the model does not only need to produce the same performance as the twinned human, but also the same sequence of actions as its twin. To address these challenges, we draw from our insights from the first experiment and implement three of the proposed approaches, namely: *(i)* partial matching, *(ii) personalisation* of the model to reflect individual differences, and *(iii) synchronisation* of the model based observed response before producing the next response.

This chapter enters into the technical details of ACT-R model development which comes with the risk of losing sight of the overarching objective. Here, we clarify how the experiments connect to the rest of the thesis. Figure 4.1 shows the basic information about the experiments and future work required to achieve a full HDT. ***Experiment 1*** applies ACT-R to an industrial assembly task to test whether the model transfers to the industrial domain. The model is run in simulation mode following algorithm 2 (section 3.4.1). It produces sequences of actions which can be compared with expected human behaviour and performances to evaluate the quality of the simulations. ***Experiment 2*** tackles the challenge of emulating actions taken by an individual. The model is run in emulation mode following algorithm 1 (section 3.4.1) using both *personalisation* and *synchronisation* techniques. The evaluation of the emulation model is performed offline, leaving integration within an HDT for future works.

To achieve a full HDT ***Future Works*** need to develop the *HDT Functions* to exploit the emulation model's outputs and generate feedback messages. For the first use case (section 3.3.2), the output needs to be relayed to collaborating CPS to enable new adaptations. For the second use case (section 3.3.2), personalised guidance needs to be generated to support the twinned human performing the assembly.

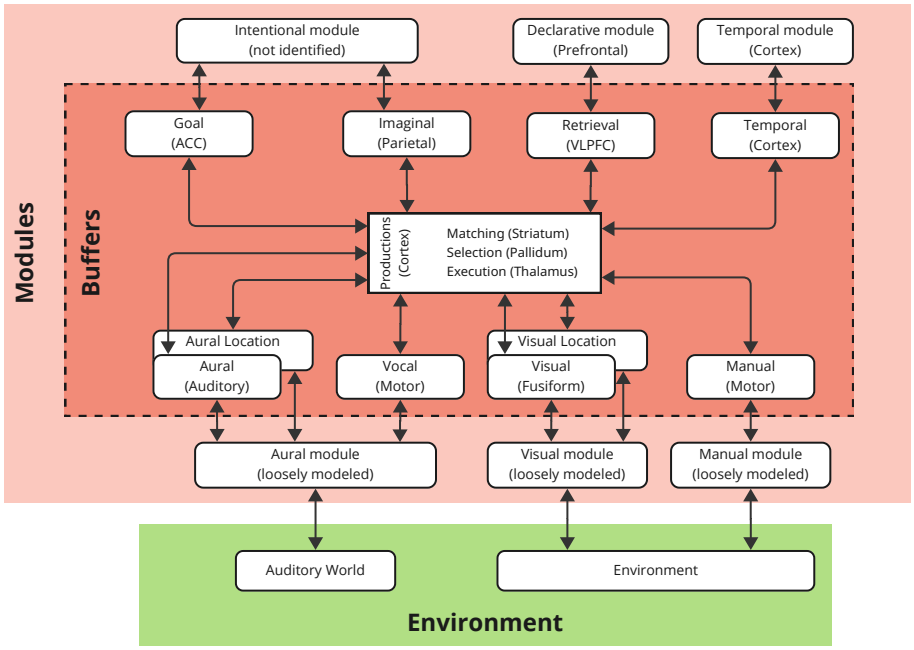


Figure 4.2: Structure of ACT-R and mapping to specific brain structures from Ritter et al. [2019]

4.1.1 Structure

The chapter is structured in the following way. First, ACT-R is described in Section 4.2. The architecture's structure and theoretical background is explained first, before diving deeper into the specific approaches and challenges to use ACT-R to model human decision making in dynamic environments. Section 4.3 maps ACT-R to our formalisation for emulation models from the previous chapter. Section 4.4 presents the first experiment to investigate the ACT-R's transferability to an assembly task. Section 4.5 presents the second experiment to evaluate the impact of *personalisation* and *synchronisation* on ACT-R's ability to emulate individual responses of a person. Finally, Section 4.6 discusses the limitations and research directions of the selected approach.

4.2 ACT-R Theoretical Background

In this section, we present the general structure and functioning of ACT-R with a deeper look at the relevant modules for our subsequent experiments. The specific modifications we apply to the architecture to adapt it to our application case are presented in the experiment sections.

4.2.1 ACT-R General Structure

ACT-R is a modular architecture, where cognition emerges from the interaction of specialised modules and buffers. Both modules and buffers are mapped to specific brain structures [Stocco et al., 2021; Ritter et al., 2019], as illustrated on Figure 4.2. The architecture is composed of independent modules, each dedicated to a specific functional domain of human cognition. These include the vision and aural modules for processing environmental stimuli, the manual (or motor) module for executing physical actions. Jointly these two modules enable the interaction between model and environment. Further internal modules are the declarative module for long-

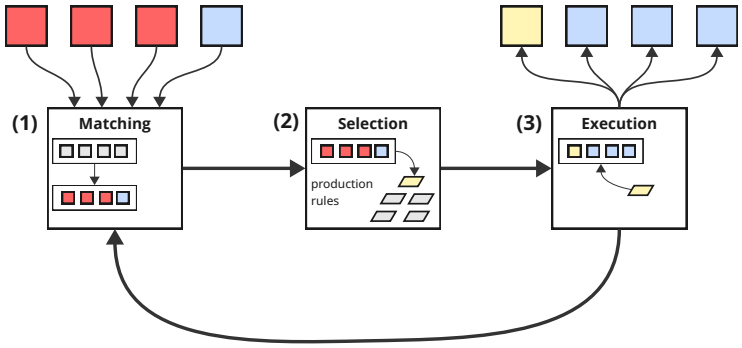


Figure 4.3: Coloured squares represent buffer states. Cognitive cycle of ACT-R in three steps **(1)** *matching* refers to reading the buffer states, **(2)** *selection* of production rule, **(3)** *execution* of production rule leading to a change in buffer state. Asynchronous execution of modules can lead to further concurrent changes in buffers.

term memory retrieval and the intentional (goal) module for maintaining task context. Lastly, the temporal module provides the architecture with time interval estimations which enables multitasking and scheduling.

The communication between the modules and central production system is mediated through buffers. Each module is associated with at least one buffer, which serves as a temporary repository for a single "chunk" of information. This architectural constrain directly represents the inherent bottlenecks and limited processing capacity of human cognition. Within this modular framework, individual modules operate in parallel and asynchronously, while the central production system coordinates their activity to produce complex, human-like behaviour.

Cognitive Cycle: The operational heart of ACT-R is the cognitive cycle, a continuous loop of matching, selection, and execution that coordinates the specialised modules to produce behaviour. As depicted on Figure 4.3, this cycle begins with the central production system examining the current content of the buffers, which hold temporary symbolic representations of the environment (perceptual stimuli) and internal mental states, such as the current task goal or a recently retrieved memory. This process is referred to as *matching*. During *selection* a production rule whose conditions match the state of these buffers is selected. Finally, the *execution* step "fires" the rule leading to discrete updates in the buffers or the initiation of new module requests. Only one production rule can be selected and executed at a time, and each firing takes a minimum of 50ms creating a bottleneck [Anderson et al., 2004]. This serial processing represents the inherent sequential nature of human decision-making and the discrete mental steps or choices a person makes when solving a task. However, while the central production system is serial, the specialised modules operate in parallel and asynchronously. For instance, a model can be shifting visual attention or executing a physical action while the declarative module is simultaneously resolving a memory retrieval request. The resulting human-like dynamics emerge from an "intricate dance" between this symbolic production system and the asynchronous modules. When the production system makes a request to a module, such as searching for an object or retrieving a fact, the module processes that request according to its own sub-symbolic mechanisms, which determine the outcome and timing based on human cognitive constraints. Once a process is resolved, the module feeds the result back into its associated buffer, where it becomes available to trigger the next cognitive cycle.

Symbolic vs. Sub-symbolic: ACT-R is a hybrid architecture that represents human knowledge through a dual-layered structure, combining discrete symbolic units with continuous sub-symbolic processes. At the symbolic level, the architecture organizes information into chunks and production rules. Chunks are the primary representations of declarative facts, goals, and sensory perceptions, while production rules encode procedural knowledge as "if-then" statements that represent the discrete mental steps or choices an individual makes during a task. This symbolic core is what makes the architecture interpretable and transparent, allowing the model's internal reasoning and goal states to be explicitly inspected.

The sub-symbolic level serves as the underlying engine that determines the dynamics and constraints of these symbolic structures using mathematical equations and quantitative metadata. While the production rules define the logical flow of a task, sub-symbolic mechanisms determine the outcome and timing of cognitive processes within the modules. For example, every chunk in declarative memory has an activation level that fluctuates based on factors such as how recently and frequently it has been reinforced through use. This sub-symbolic activation directly dictates the probability of successful retrieval and the temporal latency required to access information, enabling the model to realistically simulate human-like learning, forgetting, and confusions. This integrated logic provides a critical advantage for cognitive emulation by ensuring biological plausibility. By modulating symbolic choices with sub-symbolic constraints, such as the 50ms production firing bottleneck and memory decay, ACT-R respects the temporal limitations and bounded rationality inherent in human cognition.

Modelling approaches: Implementing a functional ACT-R model requires transitioning from architectural theory to a task-specific representation of human behaviour. This process requires the creation of production rules that accurately encode the discrete mental steps and choices a human makes when performing a specific task. Developing these rules requires expert knowledge of the task domain and is often refined through interviews and behavioural observations to ensure the model's symbolic logic aligns with the reported reality of the human worker. This hand-crafting of rules ensures the model remains transparent and interpretable, allowing for the explicit inspection of the model's goal states and reasoning processes.

Once the symbolic structure is established, the model must be calibrated through the tuning of sub-symbolic parameters. These parameters dictate the timing and outcome of cognitive processes. A critical technical constraint of ACT-R is that the architecture is not differentiable, meaning standard gradient-descent optimization algorithms, such as ADAM [Kingma and Ba, 2014], cannot be applied. Instead, modellers typically employ parameter space exploration or Bayesian likelihood inference to identify the values that best fit observed human data. The calibration phase traditionally involves running simulations and evaluating the resulting success rates and reaction times against aggregate data from human participants. Advanced methodologies may further validate the model's internal processing by comparing it against neural response data from fMRI or EEG scans [Dimov et al., 2020]. By minimizing the error between simulated and observed performance, the modeller establishes a baseline that represents average human behaviour. This calibrated model serves as the necessary starting point for more granular refinements, such as the personalisation of cognitive traits. This methodological progression ensures that the resulting Human Digital Twin is grounded in empirical reality.

4.2.2 Declarative Memory

The availability of task-relevant knowledge drives the decisions taken during execution. For this reason, the declarative memory module, which models the learning of information and its decay, is of particular interest for our experiments and for HDTs.

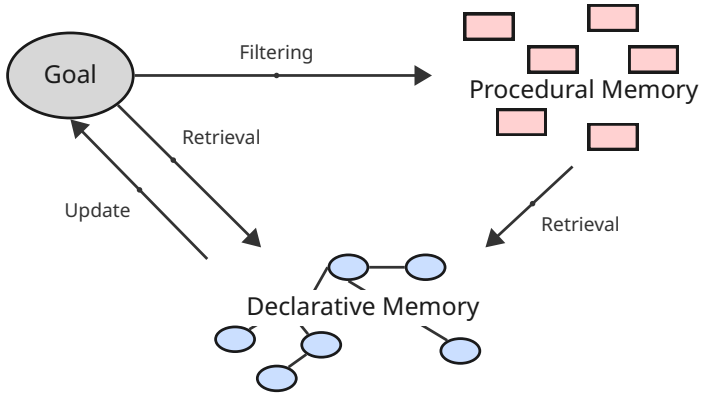


Figure 4.4: Figure showing the interactions between goal, procedural memory and declarative memory. Adapted from [Lovett et al., 1999].

In ACT-R declarative memory is represented by chunks, sometimes referred to as nodes. They can be retrieved through a retrieval request to the declarative memory module. At the moment of the request the activation A_i of each chunk i is computed to determine which chunk is selected, if any, and how long the retrieval will take. Figure 4.4 shows the interaction between the goal, procedural memory (production rules) and declarative memory. The goal acts as a filter on the available production rules in procedural memory. From the filtered subset one is selected and fired which advances the state of the model and may create a retrieval request to the declarative memory module. If the retrieval request has been made the activation will be computed and a chunk retrieved. The current goal boosts the retrieval chances of goal relevant chunks during this process. Once a chunk is retrieved it updates the goal and the cycle can continue. The interaction between goals, procedural memory and declarative memory forms the center of ACT-R models. Additional modules have been developed and implemented to represent processing of perceptual information and execution of motor actions in an integrated model for cognition, perception and action. This model started as an offshoot of ACT-R with the name ACT-R/perceptual-motor (ACT-R/PM) and was later integrated into the main ACT-R release [Byrne and Anderson, 1998, 2001]. With or without modelling perception and motor execution, the declarative memory module and retrieval dynamics remain a core driver for model behaviour and how well it reproduces human decision making. Once a retrieval request is made the sub-symbolic process unfolds by evaluating a set of equations. Below we explain ACT-R's retrieval process and its parameters [Bothell, 2020]. These parameters are in center of our subsequent attempts to personalise ACT-R models to reflect their individual cognitive capabilities and emulate their responses.

Let $\mathcal{M}$ be the set of all declarative memory chunks $\mathcal{M} = (m_1, m_2, ..., m_N)$. Consider a retrieval request made at time t. For each memory chunk m_i the total activation A_i is computed following Equation 4.1. It is composed of the base-level activation B_i, the spreading activation

S_i, the partial matching P_i, and an additional noise ϵ_i. The total activation A_i determines the possibility and probability of retrieval for the associated chunk m_i.

$$A_i(t) = B_i(t) + S_i + P_i + \epsilon_i \tag{4.1}$$

The base-level activation B_i for a chunk m_i (Eq. 4.2) depends on how often and recently a chunk has been used or reinforced. ACT-R automatically reinforces chunks when a buffer is cleared, this includes the visual and retrieval buffer. Meaning that seeing or remembering information reinforces the associated memory chunk. In Equation 4.2, t is the moment the retrieval request is made, t_k is the time of k^{th} presentation of the information and d is a decay parameter. High decay parameter values lead to a rapid decrease of base-level activation. A constant offset β_i can be defined to represent fully acquired knowledge by keeping chunks highly activated at all times.

$$B_i(t) = \ln(\sum_k (t - t_k)^{-d}) + \beta_i \tag{4.2}$$

The spreading activation S_i for chunk m_i (Eq. 4.3) represents how goal-relevant chunks are easier to retrieve. This implies that not only the familiarity with a fact but also its relevance within a context influences the probability and ease of retrieval. Each buffer k has a set amount of source activation W_k it can spread to all chunks that share a matching value with the goal buffer state. The activation spread is further diluted by the fan effect [Anderson, 1974]: the more chunks match the value of the goal buffer state the lower the resulting strength of association. Equation 4.4 describes this relation, where S is the maximum strength of association and fan_j is the number of chunks that match the value in the goal buffer.

$$S_i = \sum_k \sum_j W_{kj} S_{ji} \tag{4.3}$$

$$S_{ji} = S - \ln(fan_j) \tag{4.4}$$

Figure 4.5 illustrates the activation spreading mechanism with an example. Declarative memory contains a set of facts about the location of Lego bricks. Each chunk links a location (drawer number) with the size and colour of the Lego brick it contains. The goal buffer specifies that the current goal is to find the brick with features $size = small$ and $colour = red$. Every chunk in declarative memory that contains the value *small* or *red* is now subject to activation spreading. The total activation the goal buffer can spread is W_{goal}. It is distributed over the two features *size* and *colour* while respecting the condition $W_{size} + W_{colour} = W_{goal}$. The second important factor is the number of chunks that match each feature. In the example only one chunk matches the colour value *red* and consequently the strength of association is $S_{red-k} = S - \log(1) = S$. The size value *small* is matched with two chunks leading to a smaller strength of association $S_{small-k} = S - \log(2) < S$.

Partial Matching P_i applies a penalty to the total activation when a chunk is only partially matched. Equation 4.5 describes the computation of partial matching. Here, MP is a tuneable parameter that determines the mismatch penalty. M_{ki} is the similarity between the k^{th} specified value of the retrieval request, and the value of the corresponding slot in chunk m_i. M_{ki} ranges from 0 for identical values to -1 for maximally different values. By default, ACT-R disables partial matching and only retrieves chunks that match all the values from the retrieval request. In reality humans can confuse similar chunks of information and retrieve the wrong one, which can be reproduced with the partial matching mechanism.

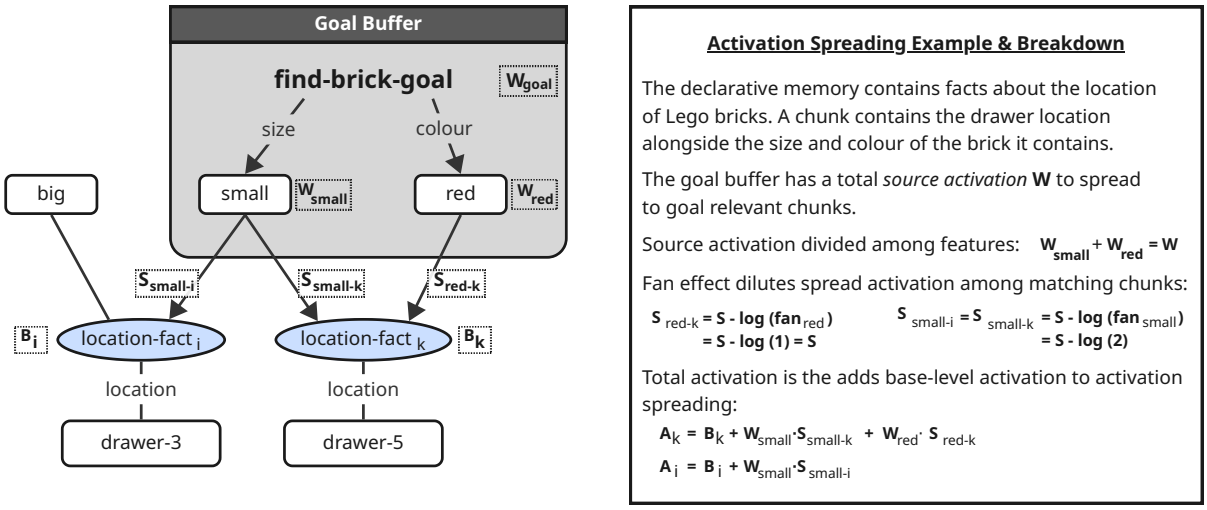


Figure 4.5: Example of activation spreading. A memory chunk encodes the location of Lego bricks that are defined by their size and colour. In the example the goal is to find the correct location of a small red brick.

$$P_i = \sum_k MP \cdot M_{ki} \tag{4.5}$$

The noise parameter ϵ_i is drawn from a logistic distribution with location parameter $\mu = 0$ and tuneable scale parameter s.

Once the total activation A_i is computed for all chunks in declarative memory $\mathcal{M}$ the chunk to be retrieved is selected. Chunks with low activations are excluded as retrieval candidates if their activation is below the ***retrieval threshold*** τ. The nodes above the thresholds are all candidates for retrieval. Recently retrieved chunks can be excluded as candidates to avoid a quick succession of retrievals and reinforcements of the same chunk. For this, the FINST mechanism tracks the recently retrieved chunks to constrain retrieval behaviour. The $FINST$ parameter defines how many chunks can be tracked simultaneously. In absence of noise the chunk with the highest activation is retrieved. If noise is applied the probability of retrieval for chunk m_i is described by equation 4.6:

$$P(\text{retrieve}_{m_i}) = \begin{cases} e^{A_i/s} / \sum_j e^{A_j/s} & \text{if } A_i > \tau \\ 0 & \text{if } A_i \leq \tau \end{cases} \tag{4.6}$$

Highly rehearsed chunks are easier and faster to retrieve, this effect is encoded in ACT-R through its retrieval latency T_i. Equation 4.7 shows the computation of the T_i, with F a tuneable time scale parameter and A_i the total activation of chunk m_i.

$$T_i = Fe^{-A_i} \tag{4.7}$$

In summary, the declarative memory module implements three major mechanisms to reproduce the reinforcement and decay of memory; ***(i)*** base-level activation models the decay of memory with time; ***(ii)*** activation spreading models the increased activation of goal-relevant information which can be interpreted as attention, and; ***(iii)*** partial matching models the possibility of confusing similar chunks of information. Each of the three mechanisms includes tuneable

parameters. Table 4.1 provides an overview of these tuneable parameters alongside a short description.

Table 4.1: ACT-R declarative module tuneable sub-symbolic parameters

Parameter	Name	Description
d	Decay rate	Memory decay as time passes based on previous reinforcements of the chunk.
β_i	Base activation	Constant activation for chunk m_i. Does not decay and can represent permanently acquired knowledge.
W_k	Activation spreading	Attentional activation to be spread by buffer k.
S	Maximum strength of association	Maximum association strength, diluted when many chunks with matching values are present.
MP	Mismatch penalty	Enables retrieval of partially matched chunks. Each mismatched slot subtracts MP from the total activation.
τ	Retrieval threshold	Only chunks with activation above this threshold can be retrieved.
F	Time scale parameter	Scales the retrieval speed of chunks.
s	Instantaneous noise	Noise is added to total activation before chunk selection and retrieval.
$FINST$ (retrieval)	declarative finsts	Avoids retrieval of same chunk in quick succession

4.2.3 Perception and Motor Modules

ACT-R's core is the interplay between the declarative memory module, production system and goal module (see Figure 4.4). Beyond the three core modules, perception and motor modules are integrated into ACT-R to model human limitations during task execution and information encoding. The modules started as an offshoot known as ACT-R/PM [Byrne and Anderson, 1998; Byrne, 2001; Byrne and Anderson, 2001] before being integrated into the main release of the architecture. The implementation is inspired by the EPIC cognitive architecture [Kieras and Meyer, 1997]. The vision module integrates the EMMA (Eye Movements and Movement of Attention) model [Salvucci, 2000] Modelling the bottlenecks of human perception, such as the visual attention or speed of encoding visual information, provides the opportunity to build cognitively plausible models with additional constrains. Taking into account the impact of the motor and perception modules is indeed one of the major modelling paradigms of ACT-R [Taatgen et al., 2006].

By default, the vision module continuously searches for new unattended stimuli but can be handled manually. Whenever a stimulus is detected the visual attention may be shifted towards it to encode the visual information at the location. The shifting of attention is controlled by two parameters; K the eye movement scaling parameter; and k the eye movement angle parameter that scales the impact of the spatial distance between two visual stimuli. Optional noise in the landing location after attention switching can be added to simulate the natural jitter of human eye movement. Additional noise can be added to the encoding duration. One last relevant

concept is visual FINST (FINgers pf INSTantiation). It marks visual stimuli as "attended" to avoid revisiting and reprocessing the same stimuli multiple times. The number of tracked stimuli can be changed. By default, up to 4 stimuli are tracked at a time, deleting the oldest one if more stimuli are processed, or decaying after a fixed time (default 3 seconds).

Table 4.2: ACT-R Vision module parameters

Parameter	Name	Description
K	Eye movement scaling parameter	Scales the overall attention switching and encoding time.
k	Eye movement angle parameter	Scales impact of distance between stimuli for attention switching and encoding.
`automatic_visual_search`	Automatic Search	Enables automatically attending to new unattended visual stimuli
$FINST$ (visual)	Visual Fingers of Instantiation	Keeps track of attended visual stimuli to avoid revisit.
`emma_landing_noise`	Landing noise	Adds noise to the landing location after visual attention switch.
`emma_noise`	Visual encoding noise	Adds noise to visual encoding time.

ACT-R's motor module simulates the execution time of keystrokes and mouse clicks. This enables the model to take into account the limitation of human execution speed. While the module is specialised on a small set of motor actions, the architecture allows for the development and integration of elaborate custom motor modules. Given that our center of interest are industrial tasks, such as assembly tasks, we opt to develop a custom motor module instead of using the standard one.

Standard perceptual and motor modules in ACT-R are traditionally optimised for computer-workstation environments, placing a specific emphasis on 2D screen-based tasks and keyboard-driven interactions. This focus has established the architecture as a mature and premier testbed for cognitive theories, exemplified by research into complex decision-making in the FireChief Microworld [Omodei and Wearing, 1995]. While modelling industrial assembly introduces requirements for 3D workspaces and intricate motor behaviours that go beyond default interactions, ACT-R's inherent modularity offers a definitive advantage. Its modular design empowers the modeller to bypass standard modules in favour of custom extensions specifically tailored to the physical demands of manufacturing. Integrating these environment-specific modules with ACT-R's biologically plausible cognitive core establishes a promising foundation for developing cognitive emulation.

4.2.4 Personalisation

Capturing individual differences is one of the five core modelling paradigms for ACT-R [Taatgen et al., 2006]. Interpersonal differences are complex and can vary in nature leading to a set of different approaches to personalise models. The main approaches are (1) sub-symbolic parameter tuning, (2) declarative memory tailoring, (3) strategic individualisation, and (4) integrative personalisation.

Sub-symbolic parameter tuning: The primary approach consists of identifying sub-symbolic parameters that correlate with the traits and abilities of individuals. These parameters

represent stable cognitive traits or biological constraints. The latency factor for retrievals F has been interpreted as an individual's tendency to respond cautiously. It explains why some individuals respond slower due to their tendency to make sure that they have the correct answer, whereas other individuals give impulsive answers [van der Velde et al., 2022]. By estimating F the authors separate the impact of caution and inherent knowledge strength on response timing. Taatgen [2002] map three parameters to established psychological constructs of the Kanfer–Ackerman Air Traffic Control task [Ackerman, 1990, 1988] for skill acquisition. Participants gain points by safely landing planes. They improve at the task through multistage learning depend on three identified parameters; (1) creating adequate mental representations requires general intelligence which was mapped to activation spreading parameter W, (2) creating mental shortcuts requires perceptual speed ability (knowledge compilation) which corresponds to proceduralisation speed, and (3) physical execution depends on psychomotor ability which is represented by the motor speed parameters. The authors succeed in reproducing patterns of correlation between individual ability and performance. Rehling et al. [2004] personalise a model's perceptual and motor parameters based on participant's mouse click speed. With the adapted parameters, the model's reaction time matches the individual's physical execution speed.

The most frequently used and researched parameter for personalisation is the source activation parameter W. It represents an individual's working memory capacity which is stable across time. It is a limited resource that makes goal-relevant information more accessible. Lovett et al. [1997, 1999] have developed an experimental protocol to empirically estimate an individual's W parameter by performing the modified digit span task. The determined parameter has been shown to transfer to other tasks [Lovett et al., 2000]. Additionally, personalising W does not only succeed on simulating an individual's retrieval success rates, but also their individual learning curves. The attempt to estimate and transfer other parameters, namely the decay rate d and retrieval threshold τ, failed to reproduce an individual's performance cross-task.

Declarative memory tailoring: This approach customises the symbolic content of a model's long-term memory to reflect an individual's experiences, knowledge and history. Marewski and Mehlhorn [2011] describe an assessment of an individual's task relevant knowledge by running a preliminary test. In this case the task performance dependent on recognising specific cities, which was tested and placed in declarative memory based on responses. The presence of chunks in declarative memory has also been derived from observed performances in a preliminary learning task [Dimov et al., 2020]. For instance, the speed of retrievals is associated with high total activation and familiarity. This can be encoded with higher base level activation to make highly familiar nodes more accessible [Taatgen et al., 2006].

Strategic individualisation: The mental processes to solve a problem differ between people among others based on their level of experience. Strategic individualisation entails developing creating symbolic production rules that represent the individual's mental process. Dimov et al. [2020] develop a set of eight different ACT-R models representing slight variations in strategies in a decision-making task where participants have to judge the best option between two choices. The utility learning mechanisms can be used to model the individual choice between competing strategies. In this case, the individual experiences a sequence of successes and failures based on their choices, which then reinforces their preferred strategy [Taatgen et al., 2006]. Multiple sets of production rules can be used to represent novice and expert strategies. How fast people change their preferred strategy depends on their individual learning rate [Stocco et al., 2024].

Integrative personalisation: Default ACT-R models do not include physiological informa-

tion as part of their processing. Morris et al. [2025] present an approach that uses information on individual circadian typologies, sleep needs and acute stress or lack of sleep to dynamically adapt model parameters. In particular, they use a fatigue module to reflect the impact of fatigue on performance. This module is not part of the standard ACT-R architecture but was developed as an offshoot [Gunzelmann et al., 2009]. The module calculates moment-by-moment fluctuations in alertness based on a person's sleep schedule and uses that information to diminish the utility of production rules in the model.

From these approaches, the potential of determining stable cognitive parameters that represent cognitive abilities of individuals are of particular interest for HDTs. They represent personalised parameters that impact central mechanisms of cognition that determine the performance and behaviour across a diverse set of tasks and situations. By contrast, implementing task specific strategies are only valid for that task and cannot transfer to other contexts. Of the parameters that have been explored for personalisation, the source activation parameter W shows the most promising to transfer across tasks.

4.3 Simulation and Emulation

ACT-R can be used to support both *simulation* and *emulation* functions of an HDT. The functioning and role of those functions are explained in detail in section 2.3.7. In section 3.4.1, we formalised the functioning of an emulation model $\mathcal{M}_\Theta$ for human cognition. It takes stimuli as an input and returns actions as an output while updating its internal state. Here, we map ACT-R to the emulation model formalisation and explain how it is used for both simulation and emulation.

In our formalisation the emulation model $\mathcal{M}_\Theta$ processes a stimulus and internal state to select an action and update the internal state (see equation 3.1). The processing is driven by a set of model parameters Θ. For ACT-R the model parameters are the sub-symbolic parameters listed in table 4.1 and table 4.2. The internal state of the model is defined by the content of its buffers and state of its modules. For our experiments the modules and their associated buffers are:

- *Vision module and buffer:* The module processes visual stimuli and places a symbolic representation of the stimuli in the associated vision buffer.
- *Declarative memory module and retrieval buffer:* The module holds all the acquired memory chunks. Through the retrieval buffer requests for retrieval can be made. The module's state is either *free*, *busy*, or *error* with the last one representing a failed retrieval attempt.
- *Goal buffer*: The goal buffer holds a symbolic representation of the current goal.
- *Motor module and buffer*: The buffer receives requests to execute motor actions that are implemented by the module. The module's state can either be *free* or *busy*.

The interaction between ACT-R and the environment happens through the perception and motor module. Figure 4.6 shows the role of the different modules during these interactions. The environment is the source of stimuli that are processed to select the next motor action in three steps: **(1)** the perception module processes and encodes the stimuli from the environment **(2)** one or multiple production rules are executed until an action is selected **(3)** the motor module executes the selected action. The processes in step 2 include the retrieval of task relevant memory

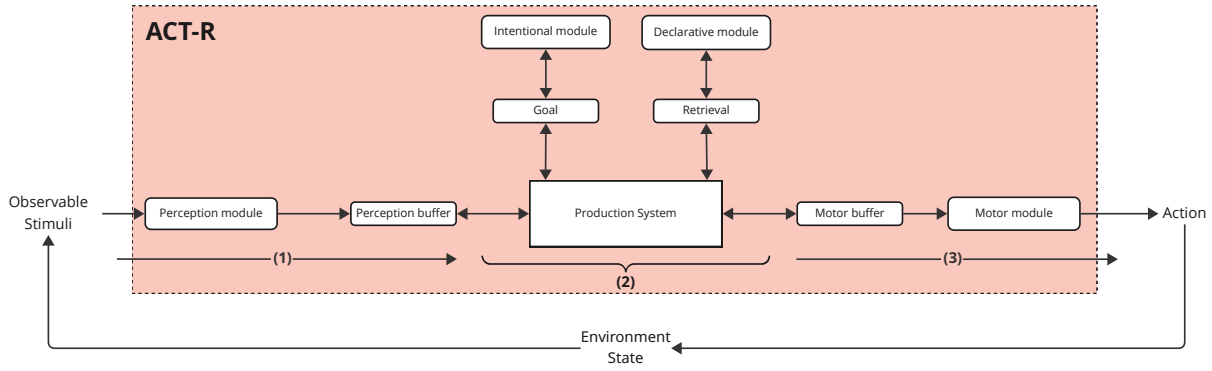


Figure 4.6: Interaction between ACT-R and environment: (1) perception of stimuli, (2) internal processing and action selection, (3) execution of motor actions.

and manipulation of the goal state. This three-step structure is aligned with the emulation model formalisation shown in Figure 3.6, with the ACT-R modules and buffer states representing the internal state from the emulation formalisation.

4.3.1 Simulation with ACT-R

ACT-R has the same input-output relationship as the formalised emulation model and can be used to run simulations following Algorithm 2 (section 3.4.1). A simulation starts from an initial state of the environment σ_0 and model s_0 and runs until it reaches the target environment state s_{target}. As an output ACT-R produces a trace that lists the sequence of all executed actions, changes of the environment state and changes of the model states with their timestamps. We use the trace to evaluate the model's ability to reproduce human behaviour and performances in two ways:

1. The model's performance during simulated tasks is analysed by extracting metrics such as success rates and error types to calibrate its sub-symbolic parameters Θ against aggregate empirical data. By minimizing the error between simulated and observed performances, this process identifies parameters that represent average human performance for the modelled task. This establishes a foundational baseline from which further refinements can be developed.

2. Simulations are used to evaluate the model's ability to reproduce individual performances by comparing simulated action traces against observed individual behaviour. This step requires the parameter tuning from the first step to further refine the model through personalization techniques (see section 4.2.4). Performance metrics, as well as exact action sequences can be used to analyse the model at a more granular level.

4.3.2 Emulation with ACT-R

For emulation ACT-R follows the algorithm 1 (section 3.4.1). Emulation starts from an initial state and is continuously synchronised with the observed actions performed by the human. At each step, ACT-R reproduces the action of the twinned human. If the action aligns with the observed action, the model can keep running. However, if the action is wrong the model needs to be synchronised with the physical reality. The core challenge is the synchronisation of the

unobservable internal state of the model based on an observable action. For the remainder of the chapter, we use the keywords emulation and simulation to refer to the two ways of running the ACT-R model:

- **Simulation** refers to ACT-R models that run from initial conditions to task completion without any synchronisation.
- **Emulation** refers to ACT-R models are synchronised after each observed response.

This corresponds to the *simulation* and *emulation* functions described in section 2.3.7. Indeed, when ACT-R runs in *simulation* mode as part of an HDT it is executed from a set of initial conditions to explore simulated human behaviour in new situations or make predictions about future decisions and states. In *emulation* mode, ACT-R is estimating the response of the twinned human given a stimulus and synchronised with the observed response, which corresponds to the emulation function of the HDT.

Note that for the subsequent experiments, ACT-R's ability to reproduce individual responses is evaluated offline using gathered observational data. This methodology is employed to quantify the impact of *personalisation* and *synchronisation* on the model's response-to-response accuracy. The model is not yet deployed within a functional HDT and does not receive real-time data; this integration work remains a key direction for future research.

4.4 Experiment 1 - Assembly Task Simulation

For this first experiment, we use ACT-R to create a simulation model for an assembly task. As previously mentioned, ACT-R has a strong track record in simulating human performances in cognitive test tasks, problem solving and skill acquisition. However, it has not been thoroughly examined for industrial assembly tasks. Our hypothesis is:

"ACT-R is able to simulate plausible human behaviour, including human error in an assembly task".

This experiment implements an ACT-R model for an assembly task where components need to be retrieved from drawers. Knowledge about the location of components represents task-relevant information that needs to be learned and can be forgotten. The behaviour resulting from the learning and forgetting is analysed and discussed to evaluate whether ACT-R can produce plausible human behaviour.

4.4.1 Task Description

The task consists of assembling 8 LEGO sets of two different types. The components are stored in a set of drawers, and the instructions are given visually on a screen. Figure 4.7(a) shows the physical setup which consists of:

- **Instruction Screen:** The screen shows visual instructions for the assembly process. The instructions are organised in a sequence of images that show the assembly steps of the product one component at a time. After the final assembly step for a product, the instructions state that a new product is starting. There are two different product types. After all 8 products are assembled, the instructions state that the task is finished.

- **Keyboard:** The keyboard interfaces with the instruction screen. To advance to the next instruction the *next* key needs to be pressed.

- **Drawers:** The drawers contain the required LEGO blocks to assemble the sets. There are 16 drawers, each containing a single type of block. Blocks come in different sizes and colours; they are either *small* (2 by 2) bricks or *big* (2 by 4) bricks; they come in 8 different colours; *red*, *yellow*, *green*, *blue*, *brown*, *black*, *grey* and *white*.

The two product types are also shown on Figure 4.7(b). Each product type is built in 9 assembly steps and requires 5 different types of blocks. For product A all the blocks are *red*, *yellow* and *green*, while product B uses *blue*, *black* and *white* blocks. Please note that product A and product B use different components without any overlap. The full task consists of assembling 8 products in a predefined sequence which is 3 times product A, then 4 times product B, and one final product A. Executing an assembly step for a product requires three steps as shown on Figure 4.7(c):

1. **Instruction step:** The instruction screen shows the starting screen or previous instruction. To start the next assembly step the instructions need to be advanced to the next slide. The information about the next assembly step needs to be encoded, including the information of what LEGO component is required to execute the assembly step.

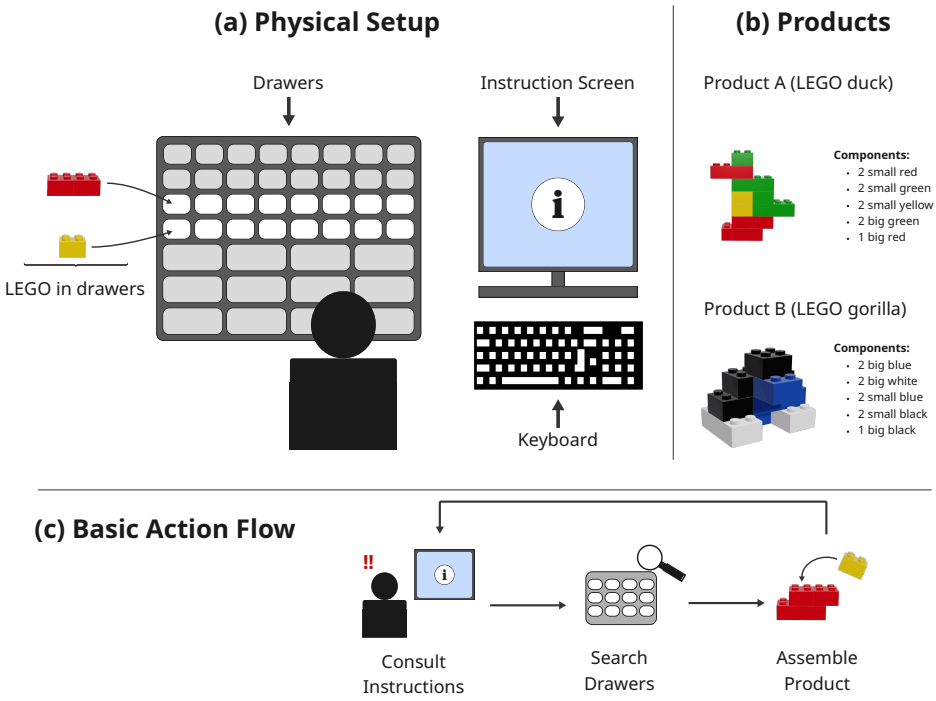


Figure 4.7: Setup for the assembly task and the two product types to assemble.

2. **Search step:** The required component is stored in one of the drawers. During the search step the drawers are explored to find the component.

3. **Assembly step:** The product is assembled using the required component in accordance with the visual instruction.

4.4.2 Model Description

To simulate the assembly task, two components are required; ***(i)*** the ACT-R model to process stimuli and select the next action and ***(ii)*** a custom environment module to resolve the selected actions, update the environment state, and provide the next stimulus. This decomposition of the overarching simulation model follows the formalisation of Figure 3.6.

Environment Module

The custom environment module allows to address the limitations of ACT-R's vision and motor module which are specialised for screen and keyboard-based tasks. The described drawer setup is 3D in nature and exceeds the scope of the provided motor and vision modules. ACT-R is a modular architecture which allows for custom modules to be developed that reflect the needs of the application case. We developed a custom environment module that tracks the environment state, handles action resolutions, and creates stimuli.

The environment state is defined by as the $\Sigma = (\Sigma_{drawer}, \Sigma_{product}, \Sigma_{screen})$. Where Σ_{drawer} encodes the state of the drawers which is defined by its content and state (open or closed). $\Sigma_{product}$ encodes the identity and state of the product that is currently being assembled. Σ_{screen}

encodes the current instruction shown on screen. It is composed of target product state of the assembly step and required component.

Each time ACT-R selects an action it places a chunk representing it into its motor buffer. From there it is resolved in two steps:

1. **Environment update:** The chunk in the motor buffer is read. The action it represents is resolved and the environment state updated to reflect its effect. This corresponds to the environment transition function from the emulation model formalisation from Section 3.4.1.

2. **Stimulus and ACT-R update:** The updated environment state may lead to a new observable stimulus. The environment module generates the stimulus and directly places it into the vision buffer. This represents the observation function from Section 3.4.1.

Table 4.3 lists all the actions and the effect of their resolution.

Table 4.3: Changes in environment module and ACT-R model when an action is resolved.

Action Type	**Environment**	**Stimulus**	**ACT-R**
Open drawer	Change drawer state to open	Drawer content becomes visible	Drawer content chunk placed in vision module
Close drawer	Change drawer state to closed	Drawer content no longer visible	No change
Take component	Updates component held in hand	No change	Goal advances to assembling the product
Assemble	Advance product state	New state of product becomes visible	Product state chunk is placed in vision module
Press next	Advance screen state	New visual instruction becomes visible	Instruction chunk is placed in vision module
Start new product	Initialise new product	Initial state of new product becomes visible	New product state chunk is placed in vision module

ACT-R Model

The ACT-R model is structured to follow the three steps of the assembly task (Instruction, Search, Assemble) as shown in Figure 4.7(c). To execute the task the model has five separate buffers and processes five chunktypes as shown on Figure 4.8. We give the details of each type in the following.

goal chunk: This chunk encodes the current goal of the model. The *goal* slot encodes in what major step the model currently is (instruction, search or assemble). The *subgoal* describes the smaller steps within the major step. The four remaining slots *size*, *color*, *location* and *product* are there to keep task relevant information accessible. Goal chunks are only accepted by the goal buffer.
instruction chunk: Instructions contain the information about the required component size and colour which are encoded in the *size* and *color* slot respectively. The target product state is encoded in the *product_ state* slot. Instructions are visual stimuli shown on screen and processed through the vision buffer..
product chunk: The identity and current state of products is encoded in this chunk. It is

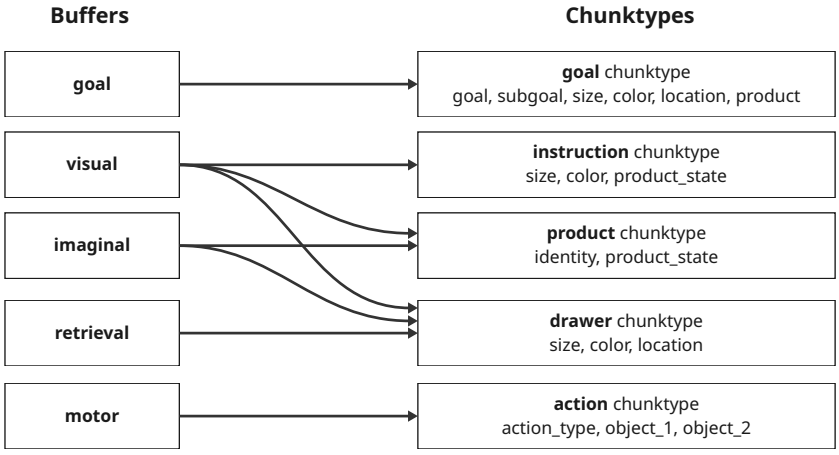


Figure 4.8: Overview of buffers and chunk types of the ACT-R model. Arrows indicate which chunktype is accepted by which buffer.

used to track the *identity* and progress of an individual product. Product chunks are processed through both the imaginal and visual buffer. Processing through the visual buffer represents seeing a product and recognising its state. When stored in the imaginal buffer, it serves as a mental representation of the current progress.
drawer chunk: Drawers are identified by their *location* and content which are components of a specific *size* and *color*. Information about drawer contents is central to the search phase. They are detected as visual stimuli and learned over time. The retrieval of drawer chunks represents recalling the location of a component.
action chunk: Selected motor actions are represented as action chunks that encode the *action_ type* as described in Table 4.3. An action is executed involving one or multiple objects that are encoded as *object_ 1* and *object_ 2*. For instance, assembling a product involves the piece held in hand and the product that is assembled.

The model's behaviour is driven by a set of production rules stored in its production system. Figure 4.9 shows the end-to-end flow of decision made by the model to execute the task. During **Instruction Step** the model is advancing the instruction shown on screen. The information is processed to determine whether the product is finished. If so, a new product is started except if it was the last product, which ends the task. From the instruction the required component is identified. The goal buffer encodes the *color* and *size* or the required piece and moves to the search step. During the **Search Step** the model makes a retrieval request through the *retrieval buffer* to recall the drawer containing the correct colour and size of component. In case of a successful recall the corresponding drawer is opened, else a random drawer is chosen. In both cases the content of the drawer is encoded in the visual buffer. If the seen content does not march the requirement a new retrieval attempt is started. This repeats until the correct component is found. The **Assembly step** is composed of three successive actions. First the component from the open drawer is taken, the drawer is closed and finally the product is assembled. Before returning the new product state is seen and encoded.

The final piece of the model is the retrieval process. Our objective is not to create optimal behaviour, but to reproduce human errors. In particular, the forgetting of task relevant information. The declarative memory module models the reinforcing and decay of memories.

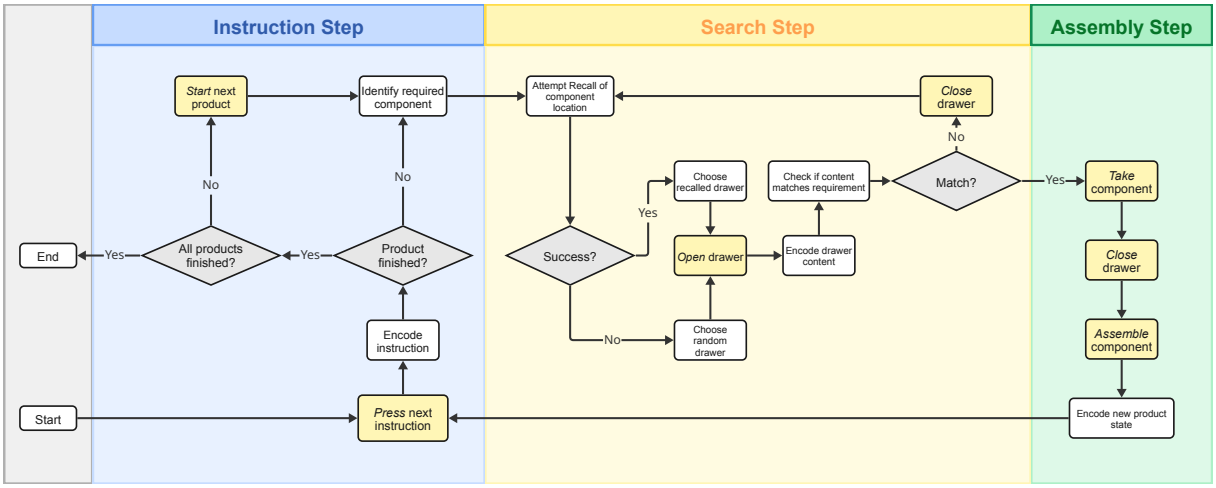


Figure 4.9: Flowchart of the model behaviour. Yellow boxes represent physical actions that change the state of the environment.

For this simulation model, we use standard parameters to analyse the model's ability to learn and forget the drawer contents. The default ACT-R model parameters for the declarative memory module use the base activation B_i computation from Equation 4.2, with decay parameter $d = 0.5$, no constant offset ($\beta_i = 0$). The default value for the decay parameter d has been extensively validated experimentally, and it is strongly recommended to maintain the standard value to reproduce plausible human forgetting behaviour [Bothell, 2020]. The default settings use no activation spreading (Equation 4.3), no partial matching (Equation 4.5) and no instantaneous noise ($s = 0$). This simplifies the computation of total activation from Equation 4.1 to $A_i = B_i$. Thus, the success of a retrieval only depends on the timing learning instances and the retrieval.

4.4.3 Simulation Results

The model is run in simulation mode as described by Algorithm 2 (section 3.4.1) to produce a model trace. The model trace reports all events produced by the ACT-R model and custom environment module. Each entry is structured as `(time, proc, action)`, where `time` is the timestamp of the event, `proc` is short for processor and refers to the module or buffer affected by the event, and `action` gives a textual description of the operation. ACT-R events are all changes to buffer states, the start of module processes and outcome of module processes. Environment events are all the changes to the environment state, in this case the resolutions of the actions.

The model succeeds in building the products using a plausible sequence of actions. It follows the three main steps of the task. During the search phase the model either recalls the location of the required component at the first try or starts randomly searching the drawers. In the latter case, the length of the search is dependent on luck during the random drawer selection.

From the model trace we extract three types performance data for the assembly task: ***(i)*** the time spent to finish the assembly of a product, ***(ii)*** the number of drawers opened to find a required component, and ***(iii)*** the success rate of the first retrieval attempt for each component. This provides a view at the granularity of the product level as well as at the level of the individual search step. Figure 4.10 shows the performance results from the simulation.

The evolution of the time spent to finish a product is shown on Figure 4.10(a). The first product, *Duck-1*, takes the longest to finish with an assembly time of 14.92s. The two next

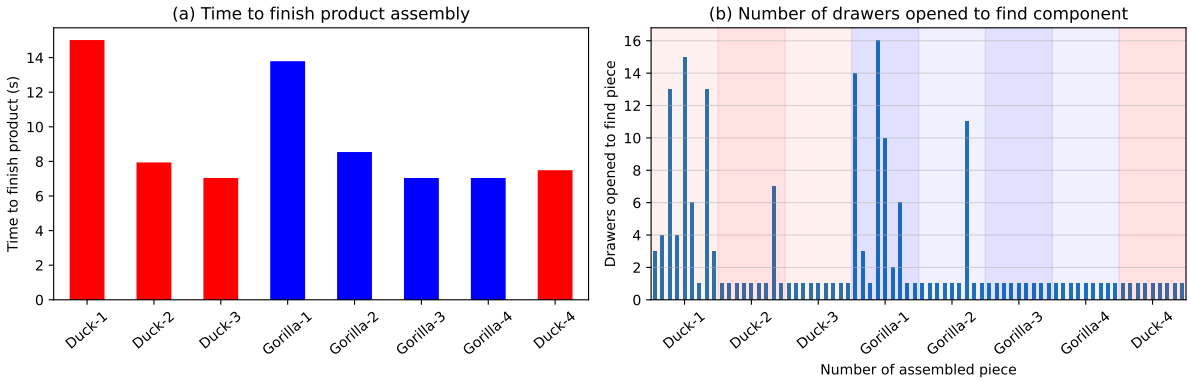


Figure 4.10: Assembly performances: (a) Time spent to finish each product. The colours indicate the product type (red for duck, blue for gorilla). (b) Number of drawers opened to find the required component for one assembly step. Each shaded area corresponds to one product.

products, *Duck-2* and *Duck-3*, show a significant speed up with a cycle time reduction of 47.2% and 53.2% respectively. This indicates that the model is learning the position of the components leading to less time spent searching for them. This is further confirmed by the number of drawers opened to find a component, as shown on Figure 4.10(b). To assemble the first product, *Duck-1*, a total of 62 drawers were opened. To find the fifth required component the model needed to search extensively, opening 15 drawers out of 16 drawers. On the second product, *Duck-2*, the number of opened drawers drops to 15 to finish the entire product. By the third product, *Duck-3*, the model finds every component at first try.

To test the model's ability to forget, we switch from assembling product A to product B. The components required to build product B are different from product A, hence they have been used and reinforced less during the first three product A assemblies. Indeed, Figure 4.10 shows an important drop in performance at the fourth product, *Gorilla-1*. To finish the product, it took 13.75s and the opening of 54 drawers. As for the previous case, the performance for the second product B assembly is already significantly faster with a 38.2% reduced cycle time. The last product if of type A. This switch back does not impact the performance; indeed every component is found on first try leading to optimal performance. Thus, the number of learning instances during the first three product assemblies were sufficient to avoid decay below the retrieval threshold.

For each assembled product, the model attempts a first retrieval of the drawer location. If it succeeds the model passes into assembly phase and takes the component. However, if it fails it can lead to a prolonged search sequence. The length of the search sequence is dependent on luck. Thus, the first retrieval success rates provide a less noisy view into the retrieval behaviour. Table 4.4 shows the success counts and success rates of the first retrieval attempts for each product. For the first product, the model lacks any prior knowledge of the drawer contents leading to a low success rate for first retrieval attempts. The success rate increases with repetitions, and the switching of product types causes a drop in performance. The switch back to product A maintains the perfect retrieval success rate.

Table 4.4: Success rate of first retrieval attempt for each product assembly.

Product Number	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
Success Count	1	8	9	3	8	9	9	9
Failure Count	8	1	0	6	1	0	0	0
Success Rate	0%	89%	100%	33%	89%	100%	100%	100%
Product Type	A	A	A	B	B	B	B	A

4.4.4 Discussion and Conclusions

The results from this experiment demonstrate that ACT-R can be used to simulate assembly tasks. Indeed, the model is capable of selecting a coherent sequence of actions to assemble the products and fulfil the production goal. In our setup, the performance is heavily dependent on the model's ability to recall the location of the components. The simulated performances show a rapid improvement after a few repetitions of assembling the same product type, demonstrating that the model is able to learn and find task relevant components faster. When switching to another product type, the performance decreases significantly which confirms that the model is also forgetting.

While the model is capable of learning and forgetting, the simulated execution times and success rates exceed human capabilities. From preliminary observations of human performance, we found that assembling a single product takes approximately 57 seconds when every component is found at first try. The simulated time to assemble a product with perfect recalls is 7s, which is over 8 times faster than our preliminary observation. This discrepancy is due to the difficulty of simulating realistic execution times of motor actions. ACT-R provides a validated motor module that can produce realistic execution times of keystrokes and mouse clicks. However, more complex actions like drawer openings and assembly are not modelled. This was the motivation to develop the custom environment module. This module is successfully resolving the actions and updating the environment state. In its current form, it is not including the simulation of realistic execution times of motor actions but resolves them instantly. To address this limitation, we see two viable approaches depending on the model execution mode. For *simulations*, the custom environment module needs to be expanded to include execution times of each action. As first approximation, a constant offset could be used for each action type, but this would ignore the impact of the physical setup and geometry influencing the motor actions. For more realistic simulations existing frameworks that model human motion like HUMOSIM could be integrated [Reed et al., 2006; Zhou et al., 2009]. For *emulations*, an alternative approach may be viable. As the model is updated after each observed action (see Algorithm 1, section 3.4.1), the timing of the observed action can be used to resynchronise the timing and avoid major drifts between the model state and observed reality. As memory decay is highly time sensitive, the synchronisation with observed timings will also impact the retrieval success and subsequent action selections. *Synchronisation* can also be used to update the internal state, in particular declarative memory, based on the observed actions.

The model's core mechanism we exploit is the retrieval of memories. The simulations have shown rapid learning and limited forgetting that exceed human capabilities. From preliminary observations, we found that reaching near perfect recalls requires 8 repetitions of the same product assembly. The model simulations achieve perfect recall within only 3 repetitions. The current ACT-R model uses default settings and excludes more advanced parts of its memory mechanisms. Indeed, we only exploit the base activation mechanisms that solely depends on time. To achieve

more plausible results, ACT-R's *partial matching* and *activation spreading* mechanisms can be used. *Partial matching* enables the model to confuse LEGO components by their colour or size. In the current implementation the model either retrieves the correct piece, or fails to retrieve any chunk, then moves to select a random drawer. However, humans don't necessarily pick a random but retrieve the incorrect memory. Partial matching allows to reproduce this behaviour, by penalising the activation of mismatched chunks instead of excluding from retrieval (see equation 4.5). *Activation spreading* models how not only the familiarity with a fact but also its relevance within a context influences the probability and ease of retrieval. This mechanism adds a time independent component to the retrieval behaviour (see equation 4.3) and represents working memory capacity. It has successfully extensively used to reproduce the performances of individuals instead of mean population performances (see section 4.2.4), by tuning the associated parameter to reproduce individual performances.

Our conclusion from this first experiment is that ACT-R is indeed able to learn and forget task relevant information during assembly tasks. However, its performance far exceeds human capabilities and require a better parametrising from the default one. To approach human performances, we identified four promising avenues we explored:

(i) *expanding timings* of the simulation of motor action;

(ii) using *partial matching* to allow for the retrieval of wrong drawers, instead of selecting randomly;

(iii) *personalisation* through activation spreading mechanism, and;

(iv) for emulation, *synchronising* the timing and internal state with the observed actions.

In the present assembly task, the central driver for task performance and action selection is the retrieval behaviour. To address the identified limitations, we present a second experiment in the following section 4.5. Its model incorporates *partial matching* to realistically account for memory confusion and retrieval errors. Discrepancies in simulated motor execution timing are addressed through two distinct ways: the application of constant offsets during simulations and the synchronisation of model time with observed action timings during emulation. The core objective of this second experiment is to evaluate how *personalisation* and *synchronisation* influence retrieval behaviour, validating these results against observed human performance to explore the limits of ACT-R for emulation purposes

In this assembly task, retrieval behaviour serves as the primary driver for both task performance and action selection. To address the limitations identified in Section 4.4.4, we present a second experiment in Section 4.5. This model incorporates partial matching to realistically account for memory confusion and retrieval errors. Discrepancies in simulated motor execution timing are addressed through two distinct methodologies: the application of constant offsets during simulations and the synchronization of model time with observed action timings during emulation. The core objective of this second experiment is to evaluate how personalization and synchronization influence retrieval behaviour, validating these results against observed human performance to explore the functional limits of ACT-R for emulation purposes

4.5 Experiment 2 - Retrieval Behaviour Emulation

Through the first experiment, we demonstrated that ACT-R is capable of simulating a coherent sequence of actions for an assembly task, reproducing human-like learning and forgetting trends. However, the initial model functioned strictly as a simulation, revealing significant limitations: the rate of learning and forgetting far exceeded human capabilities, and the lack of realistic motor timings led to cycle times over eight times faster than observed human performance. Furthermore, by relying on default sub-symbolic parameters, Experiment 1 ignored interpersonal differences, treating the worker as a generic entity rather than a unique individual.

For an HDT to effectively enable cognitive interoperability and adaptive collaboration, it requires an emulation model. Thus, ACT-R must move beyond aggregate simulation to an emulation model of its twinned individual. To achieve this, we hypothesise that the model must fulfil two critical requirements: ***(i)*** synchronisation to remain aligned with the twinned individual's state, and ***(ii)*** personalisation to reflect the individual's unique cognitive characteristics, such as working memory capacity.

In response to these challenges, Experiment 2 explicitly works towards this emulation capability by empirically evaluating individual retrieval behaviour at a response-to-response granularity. The objective of this experiment is to quantify the impact of *personalisation* and *synchronisation* on the model's accuracy during the search step of an assembly task. As illustrated in Figure 4.11, the experiment is structured into two phases:

1. **Calibration Phase:** Implements the protocol from Lovett et al. [1999] to estimate the activation spreading parameter (W) for each participant. This parameter represents the individual's unique working memory capacity and serves as the foundation for the personalised models used in the subsequent phase. Each participant performs a modified digit span task, from which the observed success rates are used to estimate their personalised W parameter.

2. **Exploration Phase:** Participants undergo a guided exploration of component locations within drawers, followed by a retrieval quiz. The resulting success rates and response sequences provide the empirical data necessary to evaluate the transition from aggregate simulation toward individual emulation. Observed success rates are used to tune the model's sub-symbolic parameters by running it in simulation mode and minimizing the discrepancy between simulated performance and observed aggregate performance. In a second analysis, we evaluate the model's ability to reproduce the sequence of individual responses using three distinct configurations to establish the impact of synchronising and personalising the model: ***(i)*** a fitted aggregate simulation model; ***(ii)*** a *personalised* model incorporating the individual W parameter; and ***(iii)*** a *synchronized* emulation model that utilizes observed responses to realign the internal cognitive state as formalized in Algorithm 1 (section 3.4.1).

By evaluating the models at this granular level, we seek to determine whether behavioural synchronisation and cognitive personalisation can effectively enable ACT-R to estimate individual responses accurately.

Sample Description We recruited participants to perform the tasks from both phases shown in Figure 4.11. Each participant is first performing the exploration task, followed by the calibration task within a single session lasting between 90 and 120 minutes. Data collection was completed first before any models were evaluated. During evaluation the personalised working

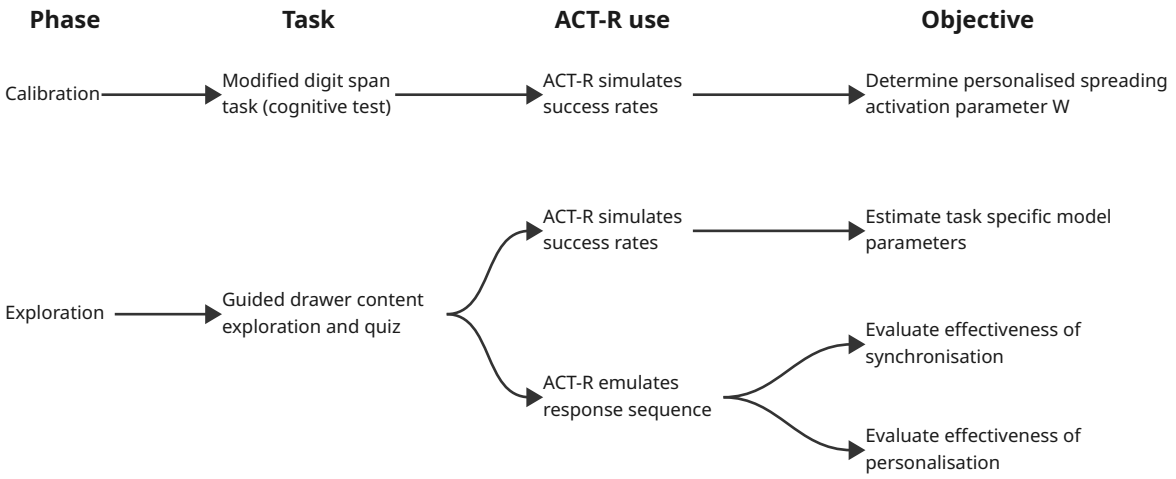


Figure 4.11: Overview of the experiment phases, tasks, ACT-R model use, and objectives.

memory parameter W is estimated based on data from the calibration phase and subsequently used in the exploration phase to personalise the models.

All participants were recruited from the Luxembourg institute of Science and Technology (LIST) on a voluntary bases, without financial or other incentives. In total 10 adult participants of working age participated in the study, 5 of them male and 5 female. Their professional backgrounds included students, engineers and scientists. We only included participants if they had no prior knowledge about the LEGO drawer task, no known cognitive deficits and no colour blindness.

We have not performed a formal power analysis to determine the sample size of the present study. As the study is exploratory in nature and with a resource-constrained sample size, future studies based on this protocol should adopt an explicit sample size justification. The observed effect size of this exploratory study may be used to perform a formal power analysis to estimate the required sample size. However, the measured effect size should be interpreted with caution as it is likely inflated due to the small sample size [Lakens, 2022].

4.5.1 Calibration Phase

The objective of this phase is to create personalised models for each participant. We implement the protocol described in Lovett et al. [1999] to estimate the working memory capacity of each participant which is represented by the spreading activation parameter W. Working memory capacity is constant in time for individuals and impacts a humans performance throughout a large variety of tasks. The method presented in Lovett et al. [1999] has been shown to transfer to other tasks [Lovett et al., 2000]. Thus, the estimates from this phase will be used to personalise the models in subsequent phases.

Note that the calibration phase does not aim to create an emulation model for the calibration task, but determines the individual working memory capacity through simulations of performances for the task.

Experimental Protocol

The protocol follows the setup and steps described in [Lovett et al., 1999]. As in the original paper, the participant sits in front of a screen where sequences of digits will be shown for a short period of time. Their task is to say every number aloud and remember the rightmost digit of each sequence. After all the sequences were displayed the participant has to recite the rightmost digits in order and without backtracking. Each trial is characterized by three parameters; ***(i)*** *number of strings* shown and thus number of digits to held in memory; ***(ii)*** the *number of digits* per string, and; ***(iii)*** *display duration* for each digit. Figure 4.12 illustrates how the digits appear on the screen with time. For each sequence, the digits appear left to right for a duration equal to the set duration parameter. The rightmost digit is displayed for twice the duration parameter.

A trial is successful if the participant has correctly read all the number aloud (Task 1) and has recalled all the rightmost numbers in the right order without back-tracking (Task 2). There are a total of 16 trial settings or types, each being repeated four times to establish a success rate. The 16 trial types are all the possible combination of the following values for the parameters; *number of strings* is $3, 4, 5$ or 6; *number of digits* per string is 4 or 6, and; *display duration* is 500ms or 700ms. The order of trial types is randomly drawn from a deck of cards.

The experiment's goal is to capture the individual differences in working memory capacity. To minimize other sources that give rise to differences of individual performance the protocol controls a few other parameters. A researcher was present during the experiment to assure that the participants stay on task for the entire duration. The tight timing constrains during the task limited the viability of diverging mental strategies for rehearsing and recalling digits. Start to finish the experiment takes approximately 50 minutes which may be long enough to induce fatigue. To mitigate the build-up of fatigue, we propose to each participant to take a short 2-minute break after 8 trial types, which is half of the experiment, is done. When the perceived difficulty increases for reasons other than task difficulty it may be an indication of fatigue, frustration or stress. At the end of the experiment, the participant is interviewed and asked about the memorization and recall strategies they deployed as well as any other factors that they believe may have impacted their performance.

Experimental Data

For each participant the success rates for all 16 trial types are recorded alongside the perceived difficulty of the task through the SEQ responses. Interview notes serve as support to identify potentially unaccounted mental strategies and identify moments of fatigue, stress and frustration. Figure 4.13 depicts the mean success rate of a dataset with 9 participants and its standard deviation. The observed performances are compared to the reported data from Lovett et al. [1999]. The general trends are the same with *number of strings* being the most impactful on difficulty and success rates. Overall, the success rates of the participants we observed are slightly more elevated especially for the long display duration of 700ms. We observe a high standard deviation that depends on the trial type. The within-task standard deviation ranger from 0.083 to 0.317 with an average of 0.21.

Model Description

We developed a model based on the available description in Lovett et al. [1999] to simulate human performances for the memory task. Since the original paper, ACT-R has evolved including a full integration of perception and motor modules that started as an offshoot (see Section 4.2.3). In 1999, ACT-R 4.0 was the newest major release and the start of integrating motor and vision

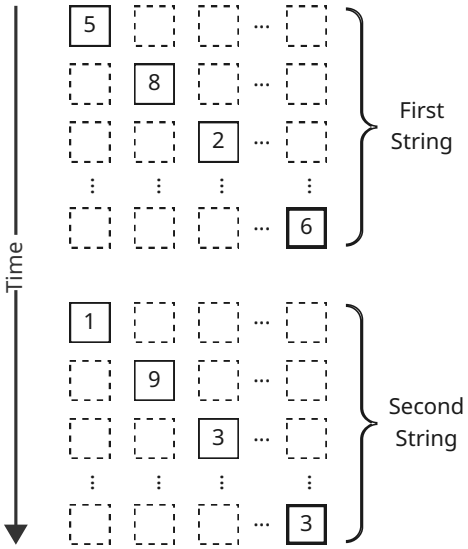


Figure 4.12: Sequence in which digits are presented on screen during the digit working memory task in the calibration phase Figure adapted from Lovett et al. [1999].

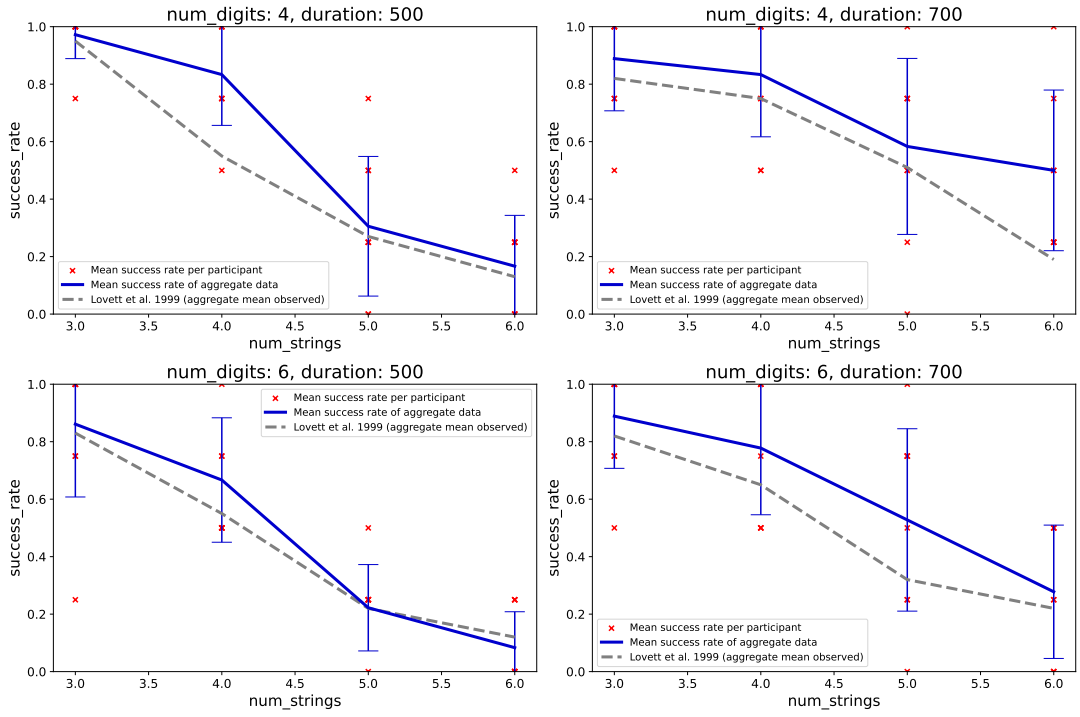


Figure 4.13: Aggregate data of observed performances compared to the original trends in Lovett et al. [1999].

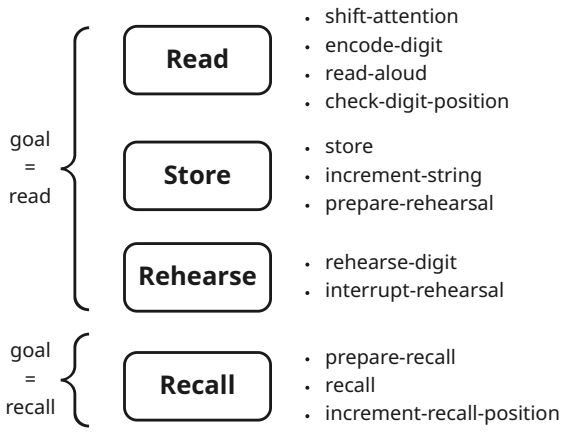


Figure 4.14: Major production rules for the calibration model and the

modules. Since then, the model has expanded in capabilities and the nuances of implementing models have changed. We are using ***pyactr***, which is an open source Python implementation of ACT-R.

The model uses ACT-R's vision module to simulate the shifting of attention and timing of encoding visual information. The main production rules of the model are the following:

- ***(i)* Read:** IF the goal is to read and digit d is on the screen THEN say digit d
- ***(ii)* Store:** IF goal is to read and digit d is on screen and d is in the last column and d has been read THEN store d and prepare to rehearse
- ***(iii)* Rehearse**: IF goal is to read and digit d is on screen and d has been read and rehearsal is prepared THEN recall stored digits
- ***(iv)* Recall:** IF goal is to recall digit in position p of trial t and digit d matches but has not been recalled THEN say digit d

The major production rules listed are in fact composed of smaller processes as shown in Figure 4.14. The goal chunk is composed a *goal*, representing the two major tasks (reading and recalling), and a *subgoal* that moves the model along the small processes. The **Read** production rule includes shifting visual attention to the digit on screen, encoding the digit and finally reading it aloud. During reading the position of the digit is checked and only the rightmost position moves the model to the **Store** production rule. This rule is straight forward and adds the digit to declarative memory and changes the sub-goal to rehearse. The **Rehearse** production rule consists of, retrieving the digit to be rehearsed, updating the position to be rehearsed. It continues until a new digit appears on screen which interrupts the ongoing rehearsal and sets the goal back to reading. Once all strings have been displayed and there is no more reading to do, the goal is set to recall. Finally, the **Recall** production rule consists of attempting retrieval of the digits in order, starting with the first remembered digit (i.e., position $p = 1$). Successfully retrieved digits are said aloud while failures lead to a *no-recall* event.

Aside from the production rules, the sub-symbolic parameters need to be defined and are subject to optimization. Table 4.5 lists the parameters of interest for the model with their name in *pyactr*, the ACT-R equivalent (see Section 4.2.2) and its value default value if applicable. The parameters that impact visual encoding are all left to their default values except all sources of additional noise have been turned off.

Table 4.5: Parameters for the calibration module.

pyactr parameter name	**ACT-R equivalent**	**Value**
decay	d	0.5 (default)
buffer_spreading_activation	W	1.0 (default) for aggregate model adapted for personalisation
strength_of_association	S	Fitted for aggregate model
retrieval_threshold	τ	0.0 (default)
instantaneous_noise	s	0.0 (default)
latency_factor	F	Fitted for aggregate model
eye_mvt_angle_parameter	k	1.0 (default)
eye_mvt_scaling_parameter	K	0.01 (default)

Experimental Results

The objective of the approach is to determine the personalised activation spreading parameter that is then used for models that model other memory-based tasks. To do so requires two steps; ***First*** the aggregate model needs to be fitted as some parameters are task dependent; ***Second*** the aggregate model is personalised by adapting only one last parameter (W) to the individual's data.

Base Model: To evaluate the effectiveness of fitting to aggregate data and the subsequent personalisation of the activation spreading parameter W, we create a Base reference model using standard values for all parameters. The only parameter that needs an estimation is the maximum strength of association S. If the value is too small, it will lead to an inhibition effect that is counter to the activation spreading mechanism as it is understood in cognitive science. An a priori estimation is $S = \log(|DM|)$ where $|DM|$ is the size of the declarative memory and log is the natural logarithm [Brasoveanu and Dotlačil, 2020]. For the digit memory task, the maximum number of chunks stored in declarative memory during a single trial is 42. For six digits per string and six strings a total of 36 digits are shown on screen and enter memory. An additional six digits are rehearsed and are stored separately. as this is the highest number of digits to remember for recall. Thus, the a priori estimate for the activation spreading parameter is $S = 3.74$. Figure 4.15 shows the simulated performances as they compare to the observed mean success rates. The base model produces high recall rates that are a poor match with the observed human performance. Indeed, the Mean Absolute Error MAE = 0.399 is elevated and the coefficient of determination $R^2 = -1.855$ is very low. This underlines the need to fit models using empirical observations as the standard parameters do not necessarily generalize well to all tasks.

Aggregate Model: The parameters of the aggregate model are fitted using the average success rates for each trial type. During optimization we seek to minimize the *Mean Absolute Error* (MAE) and use the coefficient of determination (R^2) to report performance. For each model the simulated success rates for each trial type are compared with the observed average success rates. The model's sub-symbolic processes contain stochastic elements that lead to some variance in the success rates. To evaluate the model, it is run 20 times to have a more reliable estimate of the mean behaviour and standard deviation.

In this step, we leave the majority of parameters at their default value. For the sake of completeness, we report the full parameter list in Table 4.5. Please note that the activation

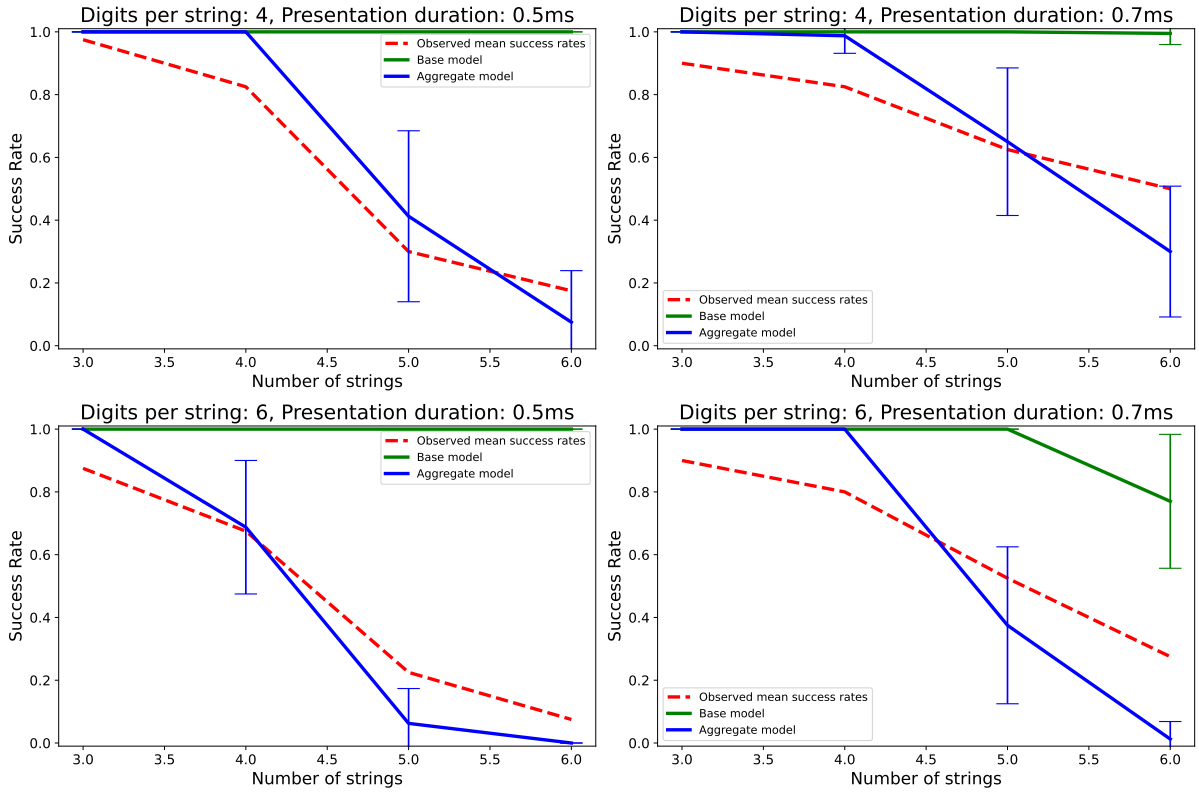


Figure 4.15: Simulated success rates compared to the observed mean success rates. The base model produces a recall success rate that exceeds human performance, while the aggregate model approaches the observed behaviour.

spreading parameter W remains at its default value $W = 1$ and will only be subject for optimization during personalisation. The two parameters that are changed to fit the model are; ***(i)*** maximum strength of association S; ***(ii)*** time scaling factor F. We perform an initial manual exploration of the parameter space to find the area that fits the observed performances best. After identifying a promising region, we do a fine-grained exploration of the area. Our best fit is for $S = 2.56$ and $F = 0.3$ with an R^2 of 0.745 and Mean Absolute Error of 0.129.

Figure 4.15 depicts the performance of the best fit. The model exhibits the same qualitative behaviour as the observations. Notably the number of strings is most impactful on difficulty and resulting success rates, while high digits per string and faster presentation rates both cause minor reductions in success rates. Both these effects are well capture by the model. In general, the model over-performs for low number of strings and declines too fast compared to observed mean success rates.

Personalised Model: With the best available aggregate model available, the next step is to personalise the activation spreading parameter W for each individual participant. The model is run for different activation spreading parameters W while maintaining all other parameters at the value of the aggregate model. For each W the model is run 20 times and its mean success rates and standard deviation per trial type are computed. To evaluate which parameter is best for an individual, we compute the MAE based on the mean success rates of each personalised model and the observed success rates of the participant. The W parameter that minimizes the MAE for an individual is retained as its personalised working memory capacity.

Table 4.6: Personalised parameters and model performances.

		MAE/R^2 for each model		
Participant ID	**Personalised W Parameter**	**Personalised**	**Aggregate**	**Base**
P01	0.825	**0.199 / 0.338**	0.220 / 0.257	0.47 / -1.972
P02	1.3	**0.134 / 0.725**	0.139 / 0.652	0.345 / -0.839
P03	1.3625	**0.089 / 0.746**	0.180 / 0.085	0.205 / -0.347
P04	0.6375	**0.081 / 0.893**	0.170 / 0.630	0.501 / -1.576
P05	1.3875	**0.125 / 0.594**	0.156 / 0.367	0.298 / -0.817
P06	0.9625	**0.128 / 0.684**	0.141 / 0.638	0.376 / -1.291
P07	0.625	**0.234 / -0.191**	0.309 / -0.728	0.595 / -4.03
P08	1.2875	**0.145 / 0.556**	0.166 / 0.415	0.313 / -0.974
P09	0.9125	**0.152 / 0.627**	0.164 / 0.594	0.485 / -1.791
P10	1.3	**0.097 / 0.81**	0.138 / 0.723	0.345 / -0.945
Aggregate	1.0	N/A	0.129 / 0.745	0.393 / -1.758

Table 4.6 summarizes the performances of all the base model, aggregate model and personalised models evaluated on individual participant data. The base model is performing poorly with a negative R^2 for every participant. Fitting the model to aggregate data is leading to a significant increase in performance across all participants. At the level of the individual the personalised model always performs best.

Discussion and Conclusion

The primary objective of this phase was to estimate the activation spreading parameter W for each participant to build personalised models. Table 4.6 shows that the personalised model are performing better than non-personalised model. The improvement is most pronounced for participants with extreme working memory capacities. Indeed, both very high $W > 1.3$ (P03, P05) and very low $W < 0.7$ (P04, P07) show the largest improvement over non-personalised models. The aggregate model uses $W = 1$. For participants with W values close to 1 the impact on performance is limited as the aggregate model already produces similar performances. We observe an anomaly with participant P07. After personalisation the performance remains very low. During the experiment they performed exceptionally well on difficult trials with 6 strings and made unexpected mistakes at easy trials with 3 strings. In the interview, the participant reports overconfidence as a reason for failing the easy trials. Additionally, the participant reported an import impact of fatigue on their performance. The ACT-R model does not include this kind of factors and is consequently unable to reproduce the individual's behaviour.

It is worth noting a few discrepancies between the results reported in the original paper [Lovett et al., 1999] and our own findings. Our aggregate model performs worse than the original paper ($R^2_{Lovett} = 0.88$ compared to $R^2_{us} = 0.745$). We have a smaller dataset with $N = 10$ with a higher variance than the original paper. Indeed, our small dataset contains both high and low performing participants. The ACT-R architecture has expanded substantially since 1999 when the reproduced personalisation approach was first published. The addition that is most impactful is the vision module based on the EMMA architecture (see Section 4.2.3). In principle, this evolution of the model should lead to better models and higher cognitive plausibility. However, it comes with an additional set of parameters to fit and sources of stochastic behaviour. EMMA is a task independent model that comes with two additional free parameters; `eye_mvt_angle_parameter` k and `eye_mvt_scaling_parameter` K. Both have come with default values, but in some cases, fine-tuning may be required. Nonetheless, the personalisation approach improves the fit between simulation and observation. For the remainder of the experiment, we use the personalised W parameter for all personalised models as listed in Table 4.6.

4.5.2 Exploration Phase

The objective of the exploration phase is to evaluate the impact of both *personalisation* and *synchronisation* on the model's ability to estimate individual responses. The participant explores the contents of a set of drawers in a guided fashion. They are instructed to focus on building up memory of the drawer contents. The experimenter guides them to open drawers one by one to learn their content and asks questions about the location of specific components to test the memory. In total 10 adult subjects participated in the study, 5 of them male and 5 female. A new set of models is developed to reproduce the success rate of recalls and the sequence of individual recall successes. The models that are differentiated by the following factors; ***(i)*** whether they are fitted to aggregate data or based on a priori estimations; ***(ii)*** whether they are personalised; ***(iii)*** whether their internal state is synchronized with observed responses. The objective of this phase is three-fold:

1. Evaluate relevance of aggregate parameter fitting: Is it necessary to fit the parameters to an aggregate population data to achieve plausible results?

2. Evaluate transferability of personalisation: Does the personalised activation spreading pa-

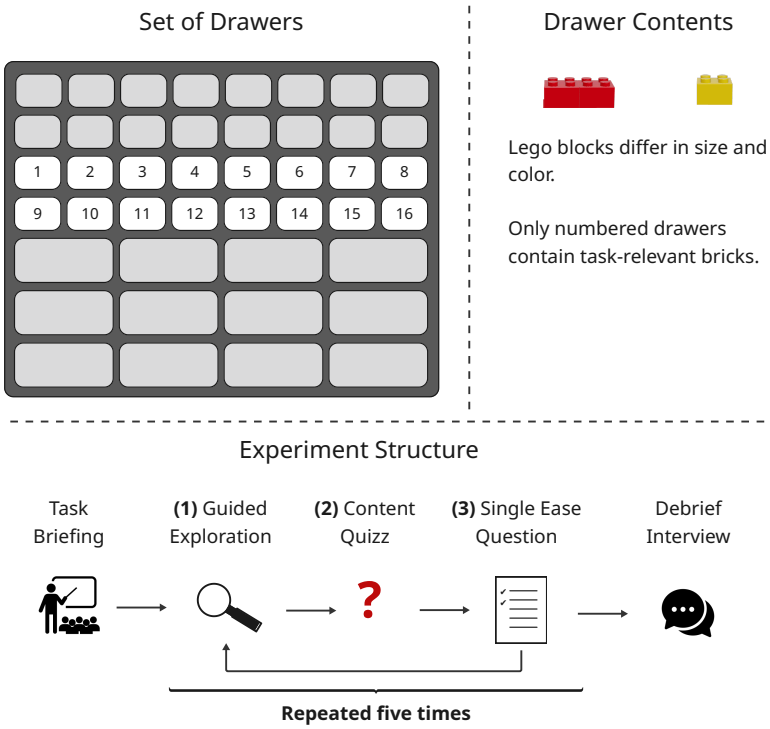


Figure 4.16: Drawer setup, drawer contents and structure of the exploration experiment.

rameter W from section 4.5.1 improve the performance of the model during the exploration task?

3. Evaluate effectiveness of synchronisation: Does synchronizing the state of the model with observed responses improve the performance?

Experimental Protocol

The participant is seated in front of a set of drawers containing different types of LEGO pieces. Figure 4.16 illustrates the drawer setup and experiment structure. In total 16 drawers are labelled with a number, running from 1 to 16. The labelled drawers contain the LEGO pieces while all other drawers contain nothing relevant to the task. LEGO pieces come in different sizes and colours; they are either *small* (2 by 2) bricks or *big* (2 by 4) bricks; they come in 8 different colours; *red*, *yellow*, *green*, *blue*, *brown*, *black*, *grey* and *white*. Each labelled drawer contains a single type of LEGO pieces. Which drawer contains what LEGO type is randomized for each participant. The drawers are equipped with sensors to automatically detect when they are opened or closed. The task is structured in three steps that are repeated five times;

1. **Exploration step:** The researcher calls out the number of a drawer; the participant opens the called out drawer, says aloud what its content is by identifying the LEGO by its size and colour, and closes the drawer; the researcher calls out the next number. The order of the called out numbers is randomized for each of the five exploration steps. Once all 16 drawers have been opened exactly once the question step begins.

2. **Question step:** The researcher asks the participant where a specific LEGO piece is; the participant tries to recall the location and response with the drawer label but does not

open the drawer; the researcher opens the drawer containing the *correct* piece. During the question step the researcher records the answers.

3. **Single ease question:** The researcher presents the single ease question (SEQ); the participant answers the question to report the perceived difficulty of performing the task. The SEQ question is formulated as follows: *"Overall how difficult or easy did you find this task?"* with a seven-point rating scale ranging from *Very Difficult* (score of 1) to *Very Easy* (score of 7). Despite its simplicity the SEQ has been shown to be at least equal or even better measure for subjective task difficulty than more complicated questionnaire [Sauro and Dumas, 2009].

Before starting, the participant is briefed about the drawer setup, the LEGO contents as well as the exploration and question step. For the question step, they are instructed to answer within 4 seconds or the answer is recorded as *no-recall*. This restriction is designed to limit the use of elaborate mental strategies, such as finding the correct location by process of elimination. When switching between steps the participant has the chance to ask clarification questions about the task. After finishing all five rounds, a short interview is conducted with the participants. They are asked to report their mental strategies for memorization and the factors of the setup that influenced their performance.

Metrics

Table 4.7 provides an overview of the metrics used to evaluate models. From each participant we gather the success rates per round, and the sequence of responses. For each round, the success rate is a value between 0 and 1 leading to five data points per participant. With a total of 10 participants, we have 50 data points to evaluate model outputs against. We use Mean Absolute Error (MAE) to evaluate a model's performance. For aggregate models the data from all participants is used for evaluation, while the personalised models are evaluated using only the data of the modelled individual.

The exact sequence of responses includes the information on whether a participant could successfully recall the location of a piece. This is what we call a retrieval success or retrieval failures. From each participant, a sequence of 80 answers is recorded, with each answer either being categorized as success or failure. Modelling this sequential data of binary entries is a binary classification problem. In the present case, the mean success rates of all participants are close to balanced with a success rate of 46.25%, however the individual participants data is less balanced ranging from 21.25% to 68.75%. Additionally early rounds have lower success rates than later rounds leading to imbalanced data on both extremes. To evaluate a binary classification models with imbalanced classes, Precision, Recall and F1-score are the appropriate metrics.

Model Description

The ACT-R model is designed to respond to the instructions of the researcher one at a time, then wait for the next instruction. The experiment revolves around estimating the successes and failures of retrieval attempts during the question step. To model the mechanism the most important input is the timing of chunk reinforcements and retrieval attempts. Given the limitations of ACT-R when it comes to the visual and motor module (see Section 4.2.3), we handle the resolution of actions and encoding of visual stimuli outside of the regular ACT-R architecture with a custom environment module. Figure 4.17 shows how both types of instructions are processed.

Table 4.7: Metrics used for exploration phase.

Metric	Purpose	Application
Mean Absolute Error (MAE)	To evaluate a model's simulated performance by comparing simulated against observed success rates.	Used for aggregate models using data from all participants, and for personalised models using only the individual's data.
Precision, Recall and F1-score	To evaluate the binary classification performance of sequential data, specifically when dealing with imbalanced classes.	Used to analyse the model's ability to accurately reproduce the exact sequence responses (categorised as retrieval success or failure) for each participant.

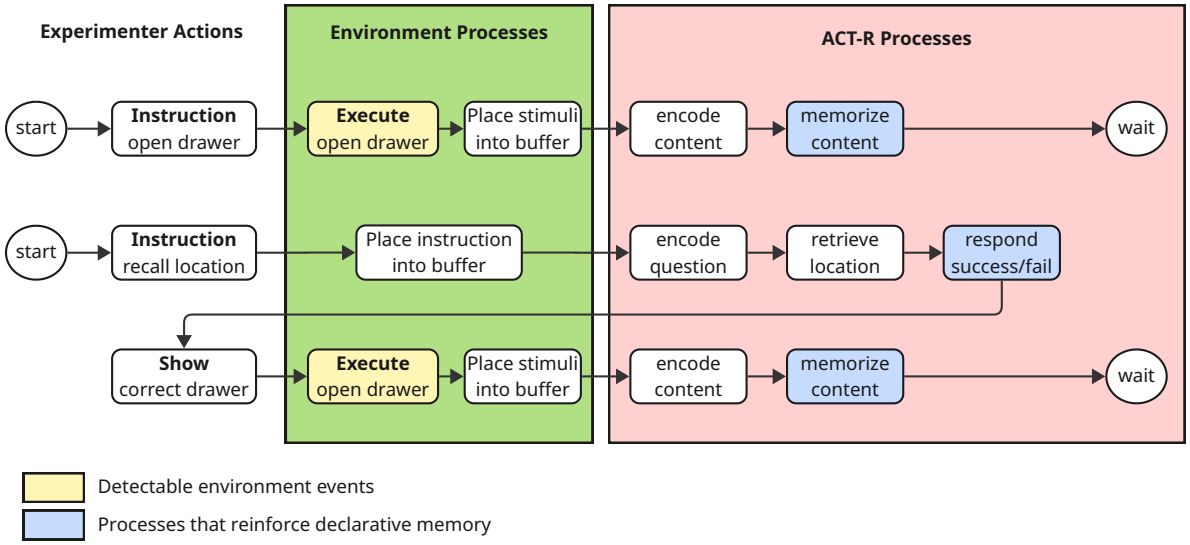


Figure 4.17: Model functioning overview. ACT-R is managing the declarative memory and retrievals. A customized Environment module is handling stimuli and executing actions.

- *Open drawer* instructions start with the experimenter verbally given the command. The environment module executes the open drawer command and places the information in ACT-R's goal buffer. Subsequently ACT-R encodes and memorizes that information. The model waits for the next instruction.
- *Recall location* instructions start with the experimenter's verbal command that states what *size* and *colour* of component to recall. The question is directly placed into ACT-R's goal buffer. ACT-R encodes the question and attempts to retrieve the location of the component. The simulated retrieval outcome (success or failure) can be compared with observed outcomes. After the response the correct drawer is opened by the experimenter to show its contents. The drawer content is encoded and memorized.

The model includes three processes that reinforce a chunk. First, for any *open drawer* instruction the content of the target drawer is seen and reinforces the chunk that links the *drawer label* to the *size* and *colour* of its content. More interesting are the reinforcements linked to the *recall location* instruction. Independent of the retrieval outcome the correct drawer is shown

and the associated chunk reinforced. The retrieval can lead to additional reinforcements, indeed there are three cases to distinguish:

- A *no-recall* event leads to no additional reinforcement
- A successful retrieval of the *correct location* reinforces the associated chunk one additional time
- A successful retrieval of the *wrong location* reinforces the incorrect chunk one additional time.

The third case is possible due to partial matching and can compound future confusions between chunks. For example, the model may retrieve and reinforce a chunk with the correct colour but wrong size. This reinforces the chunk making a future wrongful retrieval more likely. Table 4.8 summarizes the possible reinforcement depending on retrieval outcome.

Table 4.8: Reinforcements of chunks during *recall location* instructions depending on retrieval outcome.

Retrieval outcome	Correct chunk	Wrongly retrieved chunk
no recall	+1	0
retrieval of correct chunk	+2	0
retrieval of wrong chunk	+1	+1

To run simulations with the model, the instructions are given one at a time under the same conditions as in the experiment, meaning first 16 *open drawer* instructions are given, followed by 16 *recall location* instructions, with 5 repetitions. During simulation the timestamps are not available to determine the time passed between instructions. We define a constant offset between instructions of 6 seconds and a second constant offset of 60 seconds for switching from exploration step to question step and vice versa. These offsets are estimated from the observed data. It leads to a total simulated execution time of approximately 24min 30s compared to a mean observed execution time of 26min 50s.

To run emulations the model is ***synchronised*** with the observed timing of *open drawer* events and the recorded retrieval successes and failures. Each time an instruction is received a copy of the model's declarative memory is made before processing. If the estimated response is the same as the observed response the model continues without further manipulation. However, if there is a discrepancy between model estimation and observation the initial declarative memory copy is reinforced based on the observation as described in Table 4.8. The recorded timestamps associated with the observed *open drawer* events is used to set the model's internal time.

As the model does not use the standard vision module from ACT-R it does not have the same stochastic behaviour. Indeed, the model is fully deterministic if the instantaneous noise of retrieval is set to zero. The baseline sub-symbolic parameters are shown in Table 4.9.

Experimental Results

We run a set of models to test the effectiveness of our proposed *synchronisation* and *personalisation* procedures. A set of models is run to test out proposed procedures to estimate individual responses. The three approaches are; ***(i)*** creating aggregate models by fitting to population data; ***(ii)*** personalising the model based on results from calibration phase; ***(iii)*** synchronising

Table 4.9: Parameters for the exploration model. See Table 4.1 in section 4.2.2 for the full description of the parameters.

pyactr parameter name	**ACT-R equivalent**	**Value**
decay	d	0.5 (default)
buffer_spreading_activation	W	1.0 (default) for non-personalised values from calibration phase for personalised
strength_of_association	S	2.77 for base model Fitted for aggregate model
retrieval_threshold	τ	-3.68 for base model Fitted for aggregate model
mismatch_penalty	MP	1.0 (default) for base model Fitted for aggregate model
instantaneous_noise	s	0.0 (default)
latency_factor	F	0.1 (default)
retrieval.finst	FINST (declarative)	1 for all models. Default value is 0.

with observation data (see Algorithm 1, section 3.4.1). The first two approaches tune the sub-symbolic parameters Θ of the model, while the last is updating the internal state s based on observed responses. The collected data is composed of participant responses during the question rounds, SEQ question answers, and the recorded open/close drawer events. From the participant responses we extract the following information:

- The *recall success rate* for each question round and for each participant
- The sequence of successful and failed recalls, including their timing

Observed performances: Figure 4.18(a) shows the mean success rates of participants and its standard deviation. The grey crosses represent success rates of individual participants during the different question rounds. Three grey lines indicate the evolution of success rates for three participants as the rounds progress. They are a good illustration for the different learning rates individual participants experience. Notably, some participants only experience a mild increase in performance before reaching a plateau, while others keep improving until they approach near perfect success rates. This discrepancy between the participants leads to a more pronounced standard deviation in later rounds.

Figure 4.18(b) shows how the SEQ score evolves with the round number. It increases steadily as participants familiarize themselves with the task and start performing better with time. This indicates a learning effect, which is aligned with the observed performances and indicates that solid memory of the drawer setup was built during the 5 repetitions.

Impact of aggregation: We develop two models to assess the impact of fitting to aggregate data. The base model uses standard parameters with two exceptions that are a priori estimates without fitting to any empirical data. First, is the strength of association S which is estimated using the maximum size of declarative memory $S = \log(|DM|)$ [Brasoveanu and Dotlačil, 2020].

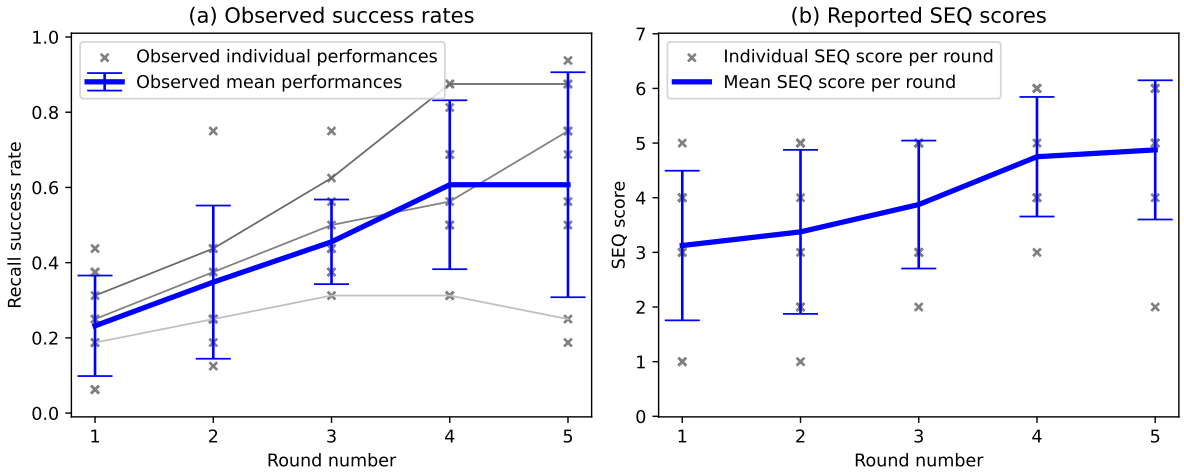


Figure 4.18: (a) Observed success rates for each round of questions. (b) Reported score on the single ease question (SEQ)

Here we have a case with a maximum of 16 chunks, one for each drawer, in declarative memory leading to $S = 2.77$. The second parameter is the retrieval threshold which is linked to the time available for retrieval. The time of retrieval T_i is computed using $T_i = Fe^{-A_i}$, where F is a constant and A_i the activation of the chunk. Enforcing a maximum retrieval time translates to setting a threshold $A_{i,min} = \tau = -\ln(T_{i,max}/F)$. During the experiment the participants had maximum 4 seconds to recall the correct drawer. With the standard value for $F = 0.1$, this leads to a a priori estimate for the retrieval threshold of $\tau = -3.68$.

The aggregate model is created by fitting τ, MP and S to the mean success rates of the participants. For the optimization we constrain the parameters around the a priori estimations to avoid implausible results. We use differential evolution algorithm because it is designed to find the global optimum without requiring a gradient [Storn and Price, 1997]. It is a population based, stochastic algorithm that maintains a random population of candidate solutions. The solutions are evaluated, mutated and recombined to find the optimal solution. For the aggregate model, the resulting optimal parameters are $\tau = -3.022$, $S = 2.45$ and $MP = 0.56$.

To evaluate the models, the same drawer configurations and instruction order was used than for the participants. For each configuration a model was executed following the associated instruction order. Each model produces a success rate for each round. The mean success rate of the model is compared with the mean success rate of the participants and evaluated using MAE. Figure 4.19 shows simulated success rates of both the base model and aggregate model. The base model's success rates are significantly higher than the observed human performance, leading to an elevated $MAE = 0.38625$. The fitted aggregate model succeeds on closely reproducing the observed performances. It has $MAE = 0.04125$. Both models have low standard deviations $\sigma_{base} = 0.0455$ and $\sigma_{aggregate} = 0.0802$ when compared to the observation's standard deviation $\sigma_{obs} = 0.1947$. For the model, the only source of this variability is the order of the instructions, indicating that the difficulty of the task is partially influenced by the order itself.

The difference in standard deviation between simulation and observation can be explained by the model's design. It simulates the behaviour and performance representing the mean cognitive capabilities of the observed population. However, it does not reproduce the extremes of high- and low-performing participants that are part of this population. To reflect these individual

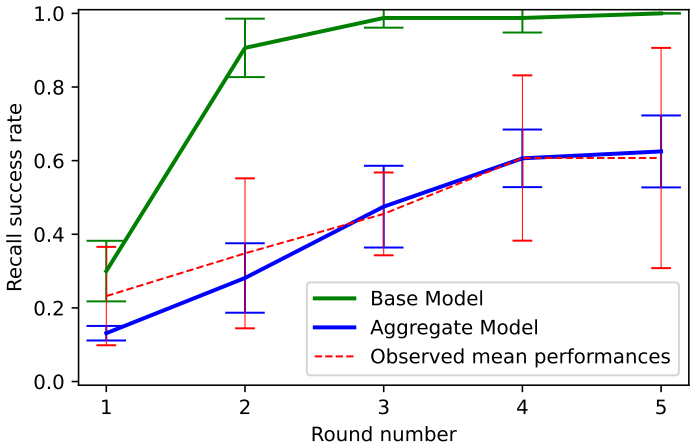


Figure 4.19: Simulated success rates of the base model (green), aggregate model (blue), in comparison with the observed mean success rate (dashed red).

differences personalised models are required.

Impact of personalisation: We create one personalised model for each individual using the estimated activation spreading parameter W from the calibration phase (see Section 4.5.1). All other parameters remain unchanged from the fitted parameters of the aggregation model. The performances of each personalised model for each individual are reported in Table 4.10. The results are mixed, and do not indicate any systematic improvement in response-to-response performance of the model. For four participants, the personalisation led to an increase in model performance. For five participants, the base model leads to better performance. Participant 06 shows no impact, this may be the result of their personalised activation spreading parameter $W = 0.9625$ being close to the value of the aggregate mode $W = 1.0$.

Table 4.10: Impact of personalisation on performance metrics.

Model	Metric	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10
regular	Precision	**0.324**	**0.225**	0.647	**0.562**	0.536	0.686	**0.531**	0.694	0.500	**0.541**
	Recall	**0.571**	0.529	0.440	**0.600**	0.375	0.436	**0.436**	0.490	0.452	**0.556**
	F1-score	**0.414**	**0.316**	0.524	**0.581**	0.441	0.533	**0.479**	0.575	0.475	**0.548**
personalised	Precision	0.278	0.214	**0.688**	0.538	**0.543**	0.686	0.529	**0.700**	**0.519**	0.404
	Recall	0.238	0.529	**0.660**	0.233	**0.625**	0.436	0.231	**0.549**	0.452	0.528
	F1-score	0.256	0.305	**0.673**	0.326	**0.581**	0.533	0.321	**0.615**	**0.483**	0.458

Figure 4.20(a) shows the change in F1-score between personalised and regular aggregate model for each participant. The x-axis is the participants personalised W parameter value, while the y-axis shows the difference in F1 score. The impact of personalisation is negative for low performing individuals, neutral for average performers and mixed for high performers. The average change in F1 score across participants is negative $\Delta F1$=-0.0335. Figure 4.20(b) shows the observed success rates during exploration for each participant and how they relate to the personalised W parameter from table 4.6. Notably, the correlation between both task performances is weak with a Pearson correlation coefficient of $r = 0.23$.

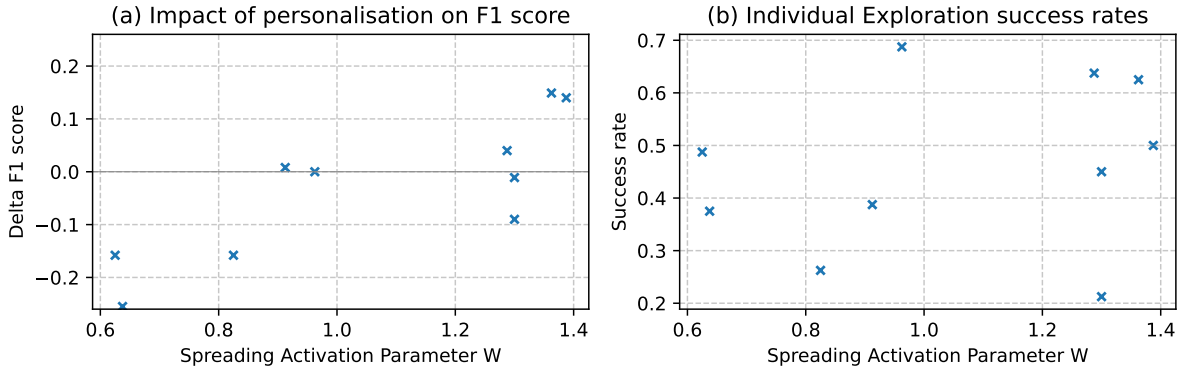


Figure 4.20: Relation of W parameters and: (a) change in F1 score between personalised and regular model, (b) observed exploration success rates.

Impact of synchronisation: The model is run using the fitted parameters of the aggregate model. Each recorded response given by the participant is used to adapt the declarative memory as described earlier. The synchronised model does not use any personalised parameters; however it follows the specific sequence of responses of the individual participant to reproduce their behaviour. Table 4.11 reports the performances of the synchronized model compared to the regular aggregate model. The synchronisation procedure leads to promising results. For every participant the recall is significantly improved. In six of the cases this comes at a minor cost of precision. The F1 score is improved in 7 cases, with 3 cases experiencing a decrease. The average impact of synchronisation across participants is a decrease of precision of -0.025, a substantial increase in recall of 0.2995, and an increase in F1 score of 0.1035.

Table 4.11: Impact of synchronisation on performance metrics

Model	Metric	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10
regular	Precision	**0.324**	**0.225**	0.647	**0.562**	0.536	0.686	**0.531**	0.694	**0.500**	**0.541**
	Recall	0.571	0.529	0.440	0.600	0.375	0.436	0.436	0.490	0.452	0.556
	F1-score	**0.414**	**0.316**	0.524	**0.581**	0.441	0.533	0.479	0.575	0.475	0.548
synchronized	Precision	0.254	0.197	**0.678**	0.383	**0.589**	**0.75**	0.508	**0.727**	0.459	0.452
	Recall	**0.762**	**0.706**	**0.800**	**0.767**	**0.825**	**0.709**	**0.846**	**0.784**	**0.903**	**0.778**
	F1-score	0.381	0.308	**0.734**	0.511	**0.688**	**0.729**	**0.635**	**0.755**	**0.609**	**0.571**

Synchronisation is effective for the majority of cases but negatively impacts the predictive power for low performing individuals. Indeed, the three cases that do not show any increase in precision and F1 score are the three cases with the lowest success rates (P01, P02, P04).

Figure 4.21(a) shows how the F1 score is evolving with passing rounds. As baseline, we use a random model whose probability to predict the positive class is equal to the prevalence of the equal class in the dataset for that round. Initial predictive power is low but increases over time for both the non-synchronised aggregate model and the synchronised emulation model. The baseline model shows increased F1 score as the prevalence of the positive class increases with the number of rounds. The aggregate model is performing at or below the level of a random predictor. However, the synchronised model improves the F1 score increasing its performance above that of a random predictor. Through synchronisation the success rates increase significantly as seen on Figure 4.21(b).

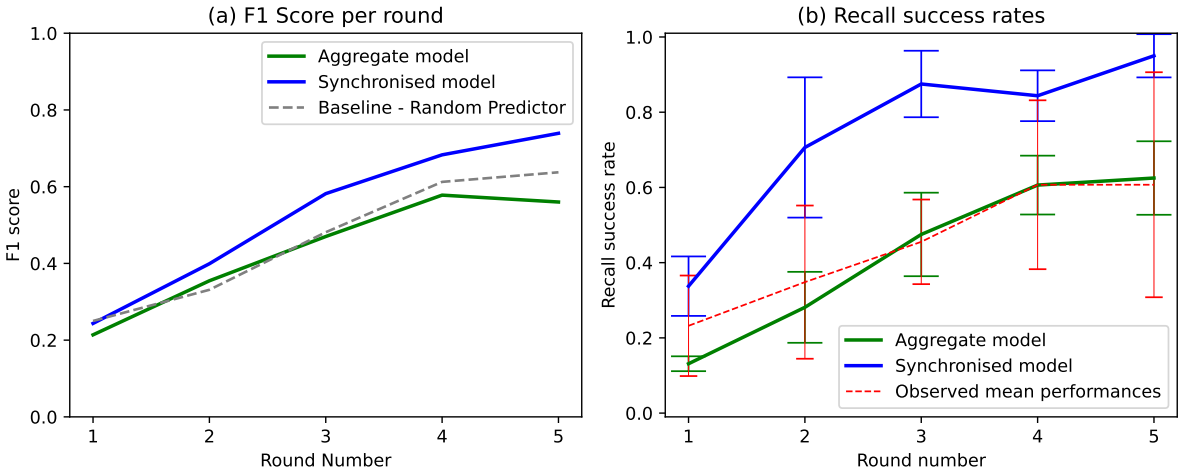


Figure 4.21: (a) F1 score for aggregate and synchronised model compared to a random predictor model. (b) Recall success rate for synchronised and aggregate model compared to observations.

Interview Results: After completing the task, participants are interviewed to elaborate on their strategy to approach the task and identify factors that may have influenced their performance. Table 4.12 summarises the self-reported strategies and influencing factors. It reveals both communalities between participants and unique approaches on the individual level.

Communalities: Colour was consistently identified as the most salient feature for memory creation, while size was frequently reported as a weak feature that leads to confusion. The most reported mental strategy is the spatial clustering of components in close proximity. Close physical contact between components, especially colour pairs significantly helped reinforcing their memory. Nearly half of participant report intuitively recalling a vague region of a component but struggling to find the exact location. This leads to near misses in their responses. The identification of these communalities can help develop a more advanced models that includes spatial reasoning and gives more weight to the colour feature to create a better general model. However, the reported interpersonal differences remain significant.

Individual differences: There are broadly two types of individual differences that were reported; *mental strategies* and the use of *prior knowledge*. While most participant relied on task-inherent cues, a small group of participant report using unrelated memories for reinforcement. Participants report that associating the drawer content or drawer number with unrelated information significantly improved memory. This includes movie titles (P06), linguistic alliterations (P08) and football match scores or jersey numbers (P03, P08). Mathematical relations were also employed by some participants to derive relative drawer locations through division or subtraction (P07, P08). Proactive learning behaviours further differentiated the group, as only a subset of participants engaged in deliberate self-quizzing during exploration (P02, P07, P10) or utilised a process of elimination to narrow down remaining choices during the question rounds (P01, P05). Finally, participants differed in their spatial methodology, with some consciously focusing on memorising the boundaries and edges of the drawers (P05, P10), while others prioritized the first row of the setup (P04).

Table 4.12: Main observations from post-task interview with participants.

Category	Description	Participant IDs
Prior Knowledge & Mnemonics	Outside-Task Associations: Linking drawer numbers or content to external information such as football match scores, player jersey numbers, or the colour of basketballs.	P03, P08
Prior Knowledge & Mnemonics	Linguistic & Aural Cues: Utilizing movie titles (e.g., The Big Blue), alliteration (shared "b" sound in French for blanc and bleu), or wordplay puns ("petit marron" to "petit marrant").	P06, P08
Prior Knowledge & Mnemonics	Mathematical Logic: Deriving relative locations through calculations, such as division by two or subtraction, to link related component pairs.	P07, P08
Mental Strategies	Clustering & Proximity: Grouping pieces that are in close physical proximity (horizontal or vertical), with a strong focus on colour pairs.	P01, P02, P03, P06, P07, P10
Mental Strategies	Edges & Extremes: Focusing first on the boundaries (edges) of the drawer setup and working inward toward the center. Or focus on the first row.	P04, P05, P10
Mental Strategies	Process of Elimination: Mentally excluding drawers that had already been visited or shown during a question round to narrow down choices.	P01, P05
Proactive Learning	Self-Quizzing: Actively attempting to predict contents or quiz oneself during exploration rounds before opening the drawer.	P02, P07, P10
Feature Salience	Identifying colour as the most reliable anchor for memory, while viewing size as a weak feature prone to confusion.	P01, P02, P03, P04, P05, P06, P07, P10
Feature Salience	Vague Region Recall: Using "fuzzy" or intuitive recall to remember a general quadrant or area rather than an exact drawer number.	P01, P02, P06, P10

Discussion and Conclusions for Experiment 2

The results of the exploration phase provide valuable insights into the challenges of emulating human behaviour at a response-to-response granularity. By comparing aggregate simulations, personalised models, and synchronised emulations, we have identified several key factors that influence the predictive power of ACT-R in industrial tasks.

The Necessity of Empirical Parameter Fitting: The first finding is that a priori parameter estimations are insufficient for realistic performance simulations. As seen in Figure 4.19, the base model using standard ACT-R parameters significantly over-performed human participants, producing near-perfect recall after only three rounds. This high learning efficiency fails to account for the natural bounded rationality of real workers. By contrast, fitting sub-symbolic parameters, specifically the retrieval threshold (τ), mismatch penalty (MP), and strength of association (S),

to aggregate empirical data resulted in a model that closely followed observed human learning trends, significantly reducing the Mean Absolute Error from 0.386 to 0.041.

Limitations of Single-Parameter Personalisation: While the activation spreading parameter (W) is a stable measure of working memory capacity [Lovett et al., 2000], the results of transferring it to the exploration task showed poor results. Personalisation improved the F1 score for high-performing individuals but had a negative impact on low performers, with an average change of $\Delta F1 = -0.0335$. The weak correlation ($r = 0.23$) between the calibration task and exploration success suggests that working memory capacity is not the sole driver of performance in assembly contexts. Instead, interpersonal differences are heavily driven by individual mental strategies and the use of prior knowledge.

The Impact of Mental Strategies and Prior Knowledge: The post-experiment interviews revealed a variety of idiosyncratic approaches participants used to improve their performance. Despite the 4 second response time limit participants report the use of approaches that go beyond direct retrieval. The findings can be organised in two categories: ***(i)*** the application of prior knowledge and ***(ii)*** the deployment of deliberate mental strategies.

Use of prior knowledge: Participants frequently utilized impactful mnemonic tools derived from knowledge outside the task context to solidify their memory of drawer contents. The identification of the prior knowledge and its impact on the performance are an important step towards improved personalised models. Some of the reported examples are hard to anticipate, such as the use of football metaphors to reinforce memory.

Mental Strategies Diverse mental strategies were employed by participants to facilitate recall despite the retrieval time limits. Such mental strategies may be implemented as a set of production rules to for example use the recall of a vague region or subsequently narrow down the candidate drawers to choose from to form a response. Self-quizzing during exploration can be reflected in the model through additional retrievals during the exploration step. However, the challenge is the differentiation between individuals that employ these strategies and those who do not.

Impact of Relaxed Protocol Timing The divergence between the model's estimations and human performance can be largely attributed to the relaxed timing of the exploration phase compared to the calibration task. While the short durations in the calibration phase left little time for rehearsal, the exploration protocol provided enough time for participants to rehearse extensively and form the elaborate associations described above. This suggests that a four-second response window may be too long to limit the viability of elaborate mental strategies. To improve the model's predictive power, these individual strategies and spatial relationships, such as the recall of a vague region, must be encoded directly into the production rules. Consequently, emulation with ACT-R may only be possible if one of the following is true: ***(i)*** The response times are short enough to limit high level reasoning. Our results demonstrate that 4 seconds are too long. This provides a first reference point to be considered by future emulation models. ***(ii)*** If response times are long, the mental strategies are known and can be encoded as production rules into the model.

Synchronisation and the Identifiability Problem: The implementation of the synchronisation procedure (Algorithm 1, section 3.4.1), yielded promising results for response-level estimations but resulted in less realistic success rates, shedding light on the inherent difficulty of inferring unobservable internal states from behavioural data. Our approach utilised the chunk

reinforcement protocol presented in Table 4.8, motivated by the fact that ACT-R naturally reinforces a chunk when it is successfully retrieved. During the simulation phase, the model frequently retrieved incorrect chunks through partial matching, reinforcing them and increasing the risk of future confusion; after parameter tuning, this implementation recreated mean observed behaviour closely.

For the emulation models, we implemented this same partial matching mechanism, reinforcing memory chunks based on the observed responses to recreate the successes and confusions of specific individuals. Additionally, the timing of responses was used to align the model's internal clock with reality, which directly impacted the amount of decay experienced by memory chunks. The increase of F1 score by 0.1035 indicates the approach is promising and partially successful. However, the sharp increase in simulated success rates (seen in Figure 4.21(b)) diverges significantly from actual human performance. We explain this discrepancy as a manifestation of the identifiability problem [Anderson, 1990]. Synchronisation relies on participant responses, yet the data shows that participants only rarely reported a "no-recall" state, preferring instead to guess blindly. Specifically, in only 12.24% of wrong answers did participants report they did not recall the location. The remaining 87.76% of wrong answers were an indistinguishable mix of genuine confusion and blind guessing. The reinforcement of chunks differs between these two mental processes:

- **Blind Guessing** does not lead to a retrieval. The participant recognises they cannot remember and most likely calls out a random drawer visible in their field of view. In this case, reinforcement should not happen.
- **Genuine Confusion:** Implies a retrieval of the wrong chunk via partial matching. In this case, reinforcement is necessary to reflect the growing strength of the incorrect association

Because our current observation data is insufficient to differentiate between these two processes, the model over-reinforced declarative memory based on guesses. While synchronisation seems to mitigate problems related to mental strategies, because the correct chunk is reinforced regardless of how the participant reached the success, the inability to distinguish a retrieval from a guess remains a hurdle in achieving high-fidelity emulation.

Summary: In conclusion, this experiment shows mixed but promising results. Our *synchronisation* protocol succeeds in improving response-by-response estimations despite the challenge of aligning internal mental states based on limited observation data. *Personalisation* based on the working memory capacity alone proved to be insufficient. To improve *personalisation* the mental strategies of individuals need to be encoded as well as prior knowledge that impacts performance.

4.6 Conclusions

This chapter focused on the third and final research question of this thesis (RQ3): *"Can cognitive architectures model human cognition and behaviour sufficiently closely to be used as emulation model for Human Digital Twins?"*. Through a two-part experimental campaign, we provided the experimental evaluation of cognitive architectures for emulation (Contribution C4), using the ACT-R architecture to model human behaviour during an industrial assembly task. The results provide a mixed answer to the suitability of cognitive architectures for HDTs. **Experiment 1** established ACT-R's transferability to the industrial domain, proving that it can simulate a coherent sequence of actions and reproduce qualitative learning trends. However, it also revealed significant limitations; the model's learning rates and motor timings initially far exceeded human capabilities. **Experiment 2** moved from aggregate simulation towards individual emulation by implementing synchronisation and personalisation. We found that:

- **Synchronisation** improved the model's accuracy (F1 score) at a response-to-response granularity, demonstrating that model state alignment is technically feasible. A major challenge is inferring cognitive states from observed actions to update the model.
- **Personalisation** using a single parameter (W) derived from a calibration task showed weak transferability. For successful personalisation individual mental strategies and prior knowledge would need to be encoded in the model.

In summary, ACT-R can probably still be considered as a promising foundation for emulating human cognition. However, it has a few important limitations that should be noted and compensated. First, model development requires hand-crafting task-specific production rules and has poor reusability. Consequently, the vision of a general model of cognition for HDTs is currently out of reach. Approaches using LLMs to generate production rules [Romero et al., 2023] or replace them [Ajraou et al., 2026] have been proposed but are still in their infancy. Second, inferring cognitive states from behavioural data is a fundamental limitation and may be limited to simple tasks that do not involve elaborate reasoning. To mitigate this limitation, richer data like reaction times or gaze tracing data could be used for the synchronisation procedure. Third, the interaction with the environment is limited with ACT-R which is particularly important to emulate physical industrial tasks. Integrating and advancing existing frameworks for embodied cognition like ACT-R/E [Trafton et al., 2013] can help, at least in theory, mitigating this limitation.

Chapter 5

Conclusions & Research Directions

This concluding chapter synthesises the findings of this thesis, summarises the contributions made toward the development of Human Digital Twins (HDTs) for Human-CPS collaboration (HCC), and outlines the future research directions required to bridge the gaps in HCC.

5.1 Thesis Summary

This thesis was conducted in the context of the AI4C2PS project, which aims to overcome the limitations of HCC by moving towards cognitive interoperability, meaning the alignment of perceptions and sharing of intentions between collaborating agents. This work's objective is to establish the foundations of HDTs and investigate their role and potential use in HCC. In particular, an HDT's role to emulate human cognition is at the center, enabling it to share valuable information about their twin's cognitive state.

The thesis progressively moves from conceptual foundations toward implementation. An in-depth systematic literature review about the conceptual fundamentals of HDTs forms the starting point of the thesis. The review exposed three major limitations: ***(i)*** There is an ambiguity about the essential elements that form an HDT, ***(ii)*** there is no unified meta-model to guide HDT development, and ***(iii)*** methods to emulate human behaviour and cognitive aspects are lacking. The limitations are addressed with a series of research questions.

RQ1: What defines a Human Digital Twin and what are its fundamental elements?

The first contribution of the thesis is the disambiguation of the *Human Digital Twin* concept through an in depth systematic literature review. We reviewed the conceptual understanding of HDTs across application domains, to extract the essential parts for HDTs. Findings were synthesised in a general domain agnostic definition and classification system, establishing the unique one-to-one relationship between the human and their HDT as well as the bi-directional automatic communication between the two. To avoid future confusion with related concepts, we propose a classification system based on level of data integration distinguishing between *Human Digital Twins*, *Shadows* and *Models*.

Building on top of the essential components identified in the definition, we develop a meta-model grounded in General Systems Theory. It clarifies the relations between the HDT, its twinned human and the systems required to enable their interaction, namely Sensors for automated data acquisition and HMI devices for feedback and manual inputs. For each system we detail the components, boundaries and interfaces for an HDT and the twinned human.

RQ2: How can HDTs be used to adapt Cyber-Physical Systems to human behaviour?

Informed by the AI4C2PS project context, we developed a decentralized Cognitive DT-based platform where every human and CPS is paired with its own digital counterpart. Within the architecture, the HDT hosts a stateful emulation model for human cognition that allows collaborating systems to draw from rich, real-time information about unobservable cognitive states, such as intentions and knowledge levels, which are traditionally undetectable to robotic sensors.

A Human-Robot Collaboration (HRC) use case illustrated how sharing estimated intent could potentially allow a robot's behavioural program to proactively avoid concurrent action conflicts and collision risks before physical movement occurs, rather than reacting to movement that is already in progress. Furthermore, the architecture could enable selective guidance, where the HDT filters environmental updates to provide the worker with only goal-relevant information or proactive warnings. This would shift collaboration from a reactive safety model to a supportive decision-making partnership that ensures shared objectives are met.

To implement this, the thesis formalised the mathematical properties of emulation models for cognition and developed an algorithm for synchronisation based on observed actions.

RQ3: Can cognitive architectures model human cognition and behaviour sufficiently closely to be used as an emulation model for Human Digital Twins?

The challenge of creating an emulation model for human cognition is the main roadblock for our proposed HDT-based adaptation approach. In the last chapter we ran two experiments to test ACT-R as potential emulation model for human cognition. From the previous HRC use case we extract the assembly task, without a robot present. We first simulate the human behaviour during the assembly task, demonstrating that the model learns and forgets task-relevant information while performing the task. In a second experiment we move towards emulation by applying two techniques; *personalisation* and *synchronisation* of the ACT-R model.

Personalisation has shown poor results, due to the limitations of encoding an individual's mental strategies and prior knowledge. *Synchronisation* with the observed actions of an individual shows promising results leading to a higher response-by-response level estimation capability

We conclude that ACT-R can still be a promising foundation to emulate human cognition. However, it faces major obstacles and limitations. First, it is unlikely to work for tasks involving higher cognition and elaborate mental strategies. Second, the models are highly task specific limiting their transferability. Third, synchronising the model's internal cognitive state based on behavioural data is inherently limited. Fourth, the current modelling of the interaction with the environment is limited and needs extension to be applicable to a wider class of industrial tasks.

5.2 Perspective and Research Directions

During this thesis, we have laid a lot of the conceptual groundwork for HDTs and their use in human-CPS collaboration. While the conceptual possibilities are promising, significant challenges remain in modelling the multifaceted nature of human beings, particularly their cognition.

A first research direction is the extension of the meta-model to provide a more granular view on all human aspects instead of focusing on cognitive aspects. Our current meta-model does not specify the interactions between various dimensions of the human. The link between physiological parameters, cognitive load, and their impact on human comfort and decision making are of particular interest for the work environment. A preliminary model identifying the main

human dimensions including their dependencies was published in Naudet et al. [2021b]. However, to achieve a complete model of the human requires continuous effort to enrich the model. Such a general model is important to move beyond task-specific HDTs that lack generalisation to broader contexts and are not designed as persistent companions to their human twin. While establishing the dependency between different human subsystems or dimensions and interlinking generalised models requires major work, it remains a key enabler for the long-term vision of HDTs: a persistent digital companion to provide support in diverse life situations

Within the scope of cognitive emulation, a wide variety of challenges and opportunities present themselves. Our findings identified critical limitations in current ACT-R implementations without excluding the architecture as a viable foundation for cognitive emulation

- While our novel approach to *synchronising* model states with observed behavioural responses shows promise, it requires further refinement. Specifically, the process of inferring unobservable cognitive states from discrete responses remains a significant hurdle. Future work should explore richer synchronization signals, such as reaction times [Dimov et al., 2020] and eye-tracking data [Balaji et al., 2023b], as avenues for achieving higher fidelity state alignment. Our experiments inferred actions from environmental changes. Incorporating granular motion-tracking data to automatically detect motor actions represents a further potential enrichment of the synchronisation signal.

- Second, the range of tasks that can currently be modelled at the response-to-response level remains restricted. High-fidelity emulation is difficult in tasks requiring long deliberation periods, as these engage higher cognitive processes and unobservable chains of reasoning. These strategies differ between individuals further complicating efforts to *personalise* models. Nevertheless, ACT-R remains promising for tasks characterised by inherent time pressure and constrained response windows [Daily et al., 2001]. A systematic identification of industrial tasks that fit this requirement would provide a valuable overview to ACT-R developers and industry practitioners.

- The impact of prior knowledge, even if not directly task-related, may impact retrieval behaviour. In that case, it must be accounted for to enable the model to reflect individual mnemonics. Current methodologies for identifying this knowledge are limited, relying primarily on preliminary interviews, tasks or quizzes.

- A further limitation is the restricted environmental interaction offered by the standard ACT-R release. For industrial and assembly tasks, a promising direction involves integrating existing skeleton and pose movement models as custom motor modules within the architecture. Similarly, perceptual modules should be extended to support interaction within 3D industrial environments. These are valuable steps towards extending ACT-R to be an embodied cognitive architecture [Trafton et al., 2013], where the interplay between human and environment directly affects cognitive processes.

Lastly, the development of ACT-R models currently necessitates the hand-crafting of task-specific production rules. This process is often cumbersome and prone to error if the underlying cognitive processes are not fully understood. While early efforts to utilize LLMs for automated rule generation [Romero et al., 2023] or replace them [Ajraou et al., 2026] have begun, this area remains largely unexplored.

For fellow researchers we provide a few recommendations based on our observations when using ACT-R with *personalisation* and *synchronisation* procedures:

1. Tasks that allow for long periods of deliberation before giving a response should be avoided, as unobserved chains of reasoning may prove impossible to encode in the model. *Personalisation* based on a single parameter may only work for response times as short as 0.9 seconds [Daily et al., 2001], after which more sophisticated approaches may be required.

2. If *personalisation* based on single parameter tuning is used (e.g., the protocol from Lovett et al. [1999]), it should be ensured that the approach transfers well to the modelled task. The correlation between performances of both tasks can provide an initial indication on whether single parameter tuning is sufficient.

3. If single parameter tuning is insufficient, the identification of individual mental strategies underlying individual differences is essential. An assessment should be made regarding whether these strategies can be encoded in the model and whether the specific strategy used by an individual can be identified through a preliminary test or questionnaire.

4. If prior knowledge is the source of individual differences, attempts should be made to identify it and evaluate whether it can be encoded in the model.

5. When implementing a *synchronisation* procedure, it is important to identify whether different mental processes lead to the same observable response. Efforts should be made to identify how the internal state of the model differs for various processes and whether additional data (such as reaction times) can help differentiate which cognitive process was used.

To bring this research full circle, we envision a multi-step progression to work towards the collaborative assembly use case (section 3.3.2). First, the advanced retrieval-focused model should be integrated into the assembly model to evaluate the impact of advanced memory mechanisms, such as partial matching and activation spreading. To apply this model within an HRC context, its scope must be expanded to account for the impact of robotic actions on human decision-making. Factors such as distractions or interruptions caused by the robot are currently not encoded in our models and represent a necessary addition. If the model successfully emulates human cognition, it can next be integrated into an HDT. All relevant actions must be made detectable in real-time to allow the model to function within an HDT and integrate with the broader AI4C2PS architecture.

This integration would represent the realisation of the systemic HDT meta-model within a wider digital environment including collaborating CPS and their DTs. Within this decentralised architecture, the HDT would share unobservable cognitive states, including intentions and knowledge levels, with CDTs of collaborating CPS. If high-fidelity state alignment is achieved between human and HDT it would provide a foundational signal to reach cognitive interoperability. It would represent a significant step towards the human-centric symbiosis envisioned by Industry 5.0, Overcoming those challenges is no possible in a short time frame, but represent a vital area for future research.

Appendix A: Systematic Literature Review Categorisation Table

In this appendix we list all the reviewed papers and in which categories they fall. We categorise papers by ***(i)*** Application Area, ***(ii)*** choice of twinned entity, ***(iii)*** level of data integration and ***(iv)*** type of study. Below we briefly explain the details of our categorisation:

Application Area: Papers are organised by application area. We identified four main application areas and use a fifth *Other* category for all areas with very few examples. The possible categories are encoded in the table as: *Industry, Healthcare, Mobility, Metaverse, Other.*

Twinned Entity: The twinned entity of an HDT is the physical entity from which data is gathered. Overall, we identified three separate categories; human individuals; human roles; systems containing humans. The possible categories are encoded in the table as: *Individual, Role, System.*

Data Integration: With the level of data integration, we distinguish papers by the types of data transfers that occur between the physical object and digital object. Digital Models (DM) have only manual data transfers between both objects. Digital Shadows (DS) have automatic data transfers from the physical object to the digital object, but the other direction is manual. Digital Twins (DT) have automated data transfers in both directions. The possible categories are encoded in the table as: *DM, DS, DT*

Study Type: We distinguish between papers that perform case studies and papers that present purely conceptual work. If a paper proposes a conceptual model and presents results from a case study in the same paper the paper will be classified as case study. Reviews are treated separately in Section 1.2 and are not included in the tables below. The possible categories are encoded in the table as: *Case-Study, Concept*

Reference	Application	Twin	Integration	Type
Barresi et al. [2022]	Healthcare	Individual	DT	Concept
Pesapane et al. [2022]	Healthcare	Individual	DT	Concept
Okegbile et al. [2023b]	Healthcare	Individual	DT	Concept
Shengli [2021a]	Healthcare	Individual	DT	Concept
Lonsdale et al. [2022]	Healthcare	Individual	DT	Concept
Iqbal et al. [2022]	Healthcare	Individual	DT	Concept
Trobinger et al. [2021]	Healthcare	Individual	DT	Case Study
De Benedictis et al. [2023]	Healthcare	Individual	DT	Case Study
Al-Zyoud et al. [2022]	Healthcare	Individual	DT	Case Study
Hernandez-Boussard et al. [2021]	Healthcare	Individual	DS	Concept
Rivera et al. [2019]	Healthcare	Individual	DS	Concept
Kobayashi et al. [2022]	Healthcare	Individual	DS	Case Study
Mourtzis et al. [2021]	Healthcare	Individual	DM	Concept
Chude-Okonkwo [2021]	Healthcare	Individual	DM	Concept
Caramihai et al. [2022]	Healthcare	Individual	DM	Concept
Jensen and Deng [2023]	Healthcare	Individual	DM	Concept
Chakshu et al. [2019]	Healthcare	Individual	DM	Case Study
Chakshu and Nithiarasu [2022]	Healthcare	Individual	DM	Case Study
Uhlenberg et al. [2023]	Healthcare	Individual	DM	Case Study
Hernigou et al. [2021]	Healthcare	Individual	DM	Case Study
Thamotharan et al. [2023]	Healthcare	Individual	DM	Case Study
Alekseyev et al. [2021]	Healthcare	Individual	DM	Case Study
Sinisi et al. [2020]	Healthcare	Individual	DM	Case Study
Tardini et al. [2022]	Healthcare	Individual	DM	Case Study
Bahrami et al. [2022]	Healthcare	Individual	DM	Case Study
Doste et al. [2022]	Healthcare	Individual	DM	Case Study
Bousdekis et al. [2020]	Industry	Individual	DT	Concept
Joseph et al. [2021]	Industry	Individual	DT	Concept
Cardin and Trentesaux [2022]	Industry	Individual	DT	Concept
Endo et al. [2023]	Industry	Individual	DT	Concept
Wang et al. [2022a]	Industry	Individual	DT	Concept
Sparrow et al. [2019]	Industry	Individual	DT	Concept
Naudet et al. [2021a]	Industry	Individual	DT	Concept
Graessler and Poehler [2017]	Industry	Individual	DT	Concept
Montini et al. [2021b]	Industry	Individual	DT	Concept
Zheng et al. [2023]	Industry	Individual	DT	Case Study
Montini et al. [2021a]	Industry	Individual	DT	Case Study
Shangguan et al. [2022]	Industry	Individual	DT	Case Study
Sparrow et al. [2022]	Industry	Individual	DT	Case Study
Löcklin et al. [2021]	Industry	Individual	DT	Case Study
Ogunseiju et al. [2021]	Industry	Individual	DT	Case Study
Zhang et al. [2023]	Industry	Individual	DT	Case Study
Liu et al. [2022]	Industry	Individual	DT	Case Study
Graessler and Poehler [2018]	Industry	Individual	DT	Case Study
Maruyama et al. [2021]	Industry	Individual	DT	Case Study
Mertens et al. [2021]	Industry	Individual	DS	Concept

Reference	Application	Twin	Integration	Type
Caroll et al. [2020]	Industry	Individual	DS	Case Study
Greco et al. [2020]	Industry	Individual	DS	Case Study
Lee et al. [2022]	Industry	Individual	DM	Case Study
Baskaran et al. [2019]	Industry	Individual	DM	Case Study
Constantinescu et al. [2019]	Industry	Individual	DM	Case Study
Sharotry et al. [2022]	Industry	Individual	DM	Case Study
Balaji et al. [2023a]	Industry	Role	DM	Case Study
Montini et al. [2022]	Industry	System	DT	Concept
Ascone and Vanderhaegen [2022]	Industry	System	DT	Concept
Donatiello and Marfia [2022]	Metaverse	Individual	DT	Concept
Wu et al. [2023]	Metaverse	Individual	DS	Case Study
Nguyen [2021]	Metaverse	Individual	DM	Concept
Jaleel [2022]	Metaverse	Individual	DM	Concept
Nguyen [2022]	Metaverse	Individual	DM	Concept
Inamura et al. [2022]	Metaverse	Individual	DM	Case Study
Stacchio et al. [2022]	Metaverse	Individual	DM	Case Study
Wang et al. [2023a]	Mobility	Individual	DT	Concept
Liao et al. [2023]	Mobility	Individual	DT	Case Study
Wang [2021]	Mobility	Individual	DT	Case Study
Wang et al. [2022b]	Mobility	Individual	DT	Case Study
Wang et al. [2023b]	Mobility	Individual	DT	Case Study
Zibuschka et al. [2020]	Other	Individual	DT	Concept
Iwaki et al. [2020]	Other	Individual	DT	Concept
Inamura [2023]	Other	Individual	DT	Concept
Furini et al. [2022]	Other	Individual	DT	Concept
Okegbile and Cai [2022]	Other	Individual	DT	Concept
Hafez [2021]	Other	Individual	DT	Concept
Martynov et al. [2022]	Other	Individual	DT	Concept
Taylor et al. [2023]	Other	Individual	DT	Concept
Chen [2021]	Other	Individual	DT	Concept
Byrne et al. [2022]	Other	Individual	DT	Case Study
Qu et al. [2024]	Other	Individual	DT	Case Study
Xiang et al. [2023]	Other	Individual	DT	Case Study
Andriopoulos et al. [2023]	Other	Individual	DT	Case Study
Gnecco et al. [2023]	Other	Individual	DS	Concept
Kopponen et al. [2022]	Other	Individual	DS	Concept
Ramos et al. [2021]	Other	Individual	DS	Case Study
Cheng et al. [2019]	Other	Individual	DS	Case Study
Boring et al. [2022]	Other	Individual	DM	Concept
Xinzhou and Kuzmichev [2022]	Other	Individual	DM	Case Study
Barricelli et al. [2020]	Other	Individual	DM	Case Study
Cheng [2022]	Other	Individual	DM	Case Study
Prokscha et al. [2022a]	Other	Role	DM	Case Study
Prokscha et al. [2022b]	Other	Role	DM	Case Study

Bibliography

Ackerman, P. L. (1988). Determinants of individual differences during skill acquisition: Cognitive abilities and information processing. *Journal of experimental psychology: General*, 117(3):288.

Ackerman, P. L. (1990). A correlational analysis of skill specificity: Learning, abilities, and individual differences. *Journal of experimental psychology: Learning, memory, and cognition*, 16(5):883.

Ahmad, R. and Plapper, P. (2016). Safe and automated assembly process using vision assisted robot manipulator. *Procedia Cirp*, 41:771–776.

Aivaliotis, P., Georgoulias, K., and Chryssolouris, G. (2019). The use of digital twin for predictive maintenance in manufacturing. *International Journal of Computer Integrated Manufacturing*, 32(11):1067–1080.

Ajraou, H., Ward, R., Farbiz, F., Graf, E., Yuan, M., Yan, S., and Oyekan, J. (2026). Act-flex: A symbolic and generative ai integration architecture for generalisable and explainable robotic disassembly tasks. *Robotics and Computer-Integrated Manufacturing*, 101:103277.

Al Haj Ali, J., Gaffinet, B., Lezoche, M., Panetto, H., and Naudet, Y. (2025a). Enabling human–cps cognitive interoperability: Cognitive architectures as technologies for human-like cognitive digital twins. *Journal of Industrial Information Integration*, page 100969.

Al Haj Ali, J., Gaffinet, B., Panetto, H., and Naudet, Y. (2024a). Cognitive systems and interoperability in the enterprise: A systematic literature review. *Annual Reviews in Control*, 57:100954.

Al Haj Ali, J., Gaffinet, B., Panetto, H., and Naudet, Y. (2025b). From semantic interoperability to cognitive interoperability: Enabling human–cps collaboration in industry 5.0. In *International Conference on Innovative Intelligent Industrial Production and Logistics*. Springer.

Al Haj Ali, J., Lezoche, M., Panetto, H., Naudet, Y., and Gaffinet, B. (2024b). Cognitive architecture for cognitive cyber-physical systems. *IFAC-PapersOnLine*, 58(19):1180–1185.

Al-Zyoud, I., Laamarti, F., Ma, X., Tobón, D., and El Saddik, A. (2022). Towards a Machine Learning-Based Digital Twin for Non-Invasive Human Bio-Signal Fusion. *Sensors*, 22(24):9747.

AlAmir, M. and AlGhamdi, M. (2022). The role of generative adversarial network in medical image analysis: An in-depth survey. *ACM Computing Surveys*, 55(5):1–36.

Alekseyev, V. V., Vizgirda, A. S., Nefedyev, D. I., and Tsareva, A. V. (2021). Measuring systems for monitoring the human state: Human digital twins based on a kinematic portrait. *Journal of Physics: Conference Series*, 1889(5):052029.

Ali, J. A. H., Gaffinet, B., Panetto, H., and Naudet, Y. (2024). Cognitive systems and interoperability in the enterprise: A systematic literature review. *Annual Reviews in Control*, 57:100954.

Allen, B. D. (2021). Digital twins and living models at nasa. In *Digital Twin Summit*.

Anastasiou, D., Gaffinet, B., and Naudet, Y. (2025). A cognitive social robot in manufacturing. In *International Conference on Social Robotics*, pages 547–552. Springer.

Anderson, J. R. (1974). Retrieval of propositional information from long-term memory. *Cognitive psychology*, 6(4):451–474.

Anderson, J. R. (1990). *The adaptive character of thought*. Psychology Press.

Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., and Qin, Y. (2004). An integrated theory of the mind. *Psychological review*, 111(4):1036.

Anderson, J. R. and Reder, L. M. (1999). The fan effect: New results and new theories. *Journal of Experimental Psychology: General*, 128(2):186.

Andriopoulos, N., Plakas, K., Mountzouris, C., Gialelis, J., Birbas, A., Karatzas, S., and Papalexopoulos, A. (2023). Local Energy Market-Consumer Digital Twin Coordination for Optimal Energy Price Discovery under Thermal Comfort Constraints. *Applied Sciences*, 13(3):1798.

Anton, F. D., Anton, S., and Borangiu, T. (2012). Human-robot natural interaction and collision avoidance in flexible manufacturing cells. *IFAC Proceedings Volumes*, 45(6):835–840.

Arai, T., Kato, R., and Fujita, M. (2010). Assessment of operator stress induced by robot collaboration in assembly. *CIRP annals*, 59(1):5–8.

Asad, U., Khan, M., Khalid, A., and Lughmani, W. A. (2023). Human-Centric Digital Twins in Industry: A Comprehensive Review of Enabling Technologies and Implementation Strategies. *Sensors*, 23(8):3938.

Ascone, C. and Vanderhaegen, F. (2022). Towards a Holistic Framework for Digital Twins of Human-Machine Systems. In *15th IFAC Symposium on Analysis, Design and Evaluation of Human Machine Systems HMS 2022*, volume 55, pages 67–72.

Asif, S., Callari, T. C., Khan, F., Eimontaite, I., Hubbard, E.-M., Bahraini, M. S., Webb, P., and Lohse, N. (2026). Exploring tasks and challenges in human-robot collaborative systems: A review. *Robotics and Computer-Integrated Manufacturing*, 97:103102.

Baheti, R. and Gill, H. (2011). Cyber-physical systems. *The impact of control technology*, 12(1):161–166.

Bahrami, F., Rossi, R. M., and Defraeye, T. (2022). Predicting transdermal fentanyl delivery using physics-based simulations for tailored therapy based on the age. *Drug Delivery*, 29(1):950–969.

Balaji, B., Shahab, M. A., Srinivasan, B., and Srinivasan, R. (2023a). ACT-R based human digital twin to enhance operators' performance in process industries. *Frontiers in Human Neuroscience*, 17:1038060.

Balaji, B., Shahab, M. A., Srinivasan, B., and Srinivasan, R. (2023b). Act-r based human digital twin to enhance operators' performance in process industries. *Frontiers in human neuroscience*, 17:1038060.

Baratta, A., Cimino, A., Longo, F., and Nicoletti, L. (2024). Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments. *Computers & Industrial Engineering*, 187:109764.

Barresi, G., Pacchierotti, C., Laffranchi, M., and De Michieli, L. (2022). Beyond Digital Twins: Phygital Twins for Neuroergonomics in Human-Robot Interaction. *Frontiers in Neurorobotics*, 16:913605.

Barricelli, B. R., Casiraghi, E., Gliozzo, J., Petrini, A., and Valtolina, S. (2020). Human Digital Twin for Fitness Management. *IEEE Access*, 8:26637–26664.

Baskaran, S., Niaki, F. A., Tomaszewski, M., Gill, J. S., Chen, Y., Jia, Y., Mears, L., and Krovi, V. (2019). Digital Human and Robot Simulation in Automotive Assembly using Siemens Process Simulate: A Feasibility Study. *Procedia Manufacturing*, 34:986–994.

Bhattacharya, S., Gaffinet, B., Naudet, Y., and Panetto, H. (2026). (in print) a formal framework for evaluating cognitive emulation models in human digital twins. *IFAC-PapersOnLine*. 23rd IFAC World Congress.

Boring, R., Ulrich, T., Park, J., Heo, Y., and Ahn, J. (2022). The hunter dynamic human reliability analysis tool: overview of the enhanced framework for modeling human digital twins. In *Proceedings of the Probabilistic Safety Assessment and Management (PSAM 16) Conference*.

Bothell, D. (2020). Act-r 7.21+ reference manual. *Pittsburgh (PA): Carnegie Mellon University*.

Bousdekis, A., Apostolou, D., and Mentzas, G. (2020). A human cyber physical system framework for operator 4.0 – artificial intelligence symbiosis. *Manufacturing Letters*, 25:10–15.

Brasoveanu, A. and Dotlačil, J. (2020). *Computational cognitive modeling and linguistic theory*. Springer Nature.

Breque, M., De Nul, L., and Petridis, A. (2021). *Industry 5.0 – Towards a sustainable, human-centric and resilient European industry*. Publications Office of the European Union.

Byrne, M. D. (2001). Act-r/pm and menu selection: Applying a cognitive architecture to hci. *International Journal of Human-Computer Studies*, 55(1):41–84.

Byrne, M. D. and Anderson, J. R. (1998). Perception and action. In *The atomic components of thought*, pages 167–200. Psychology Press.

Byrne, M. D. and Anderson, J. R. (2001). Serial modules in parallel: the psychological refractory period and perfect time-sharing. *Psychological review*, 108(4):847.

Byrne, N., Chassiakos, A., Karatzas, S., Sweeney, D., Lazari, V., Karameros, A., Tardioli, G., and Cabrera, A. G. (2022). A Multi-Level Digital Twin for Optimising Demand Response at the Local Level without Compromising the Well-being of Consumers. In *The 9th International Conference on Construction Engineering and Project Management*.

Caramihai, S. I., Dumitrache, I., Moisescu, M. A., and Sacala, I. S. (2022). Decision Support Collaborative Platform for e-Health Integration in Smart Communities Context. *Procedia Computer Science*, 214:1152–1159.

Cardin, O. and Trentesaux, D. (2022). Design and use of human operator digital twins in industrial cyber-physical systems: Ethical implications. *IFAC-PapersOnLine*, 55(2):360–365. 14th IFAC Workshop on Intelligent Manufacturing Systems IMS 2022.

Carissoli, C., Negri, L., Bassi, M., Storm, F. A., and Delle Fave, A. (2024). Mental workload and human-robot interaction in collaborative tasks: A scoping review. *International Journal of Human–Computer Interaction*, 40(20):6458–6477.

Caroll, T., Hernandez, G., Koutitas, G., Wierschem, D., Mendez, F., Vallez, D., Aslan, S., Koldenhoven, R., and Jimenez, J. (2020). Comparison of Inverse Kinematics Algorithms for Digital Twin Industry 4.0 Applications. In *2020 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*, pages 3319–3326, Toronto, ON, Canada. IEEE.

Chakshu, N. K., Carson, J., Sazonov, I., and Nithiarasu, P. (2019). A semi-active human digital twin model for detecting severity of carotid stenoses from head vibration—A coupled computational mechanics and computer vision method. *International Journal for Numerical Methods in Biomedical Engineering*, 35(5):e3180.

Chakshu, N. K. and Nithiarasu, P. (2022). An AI based digital-twin for prioritising pneumonia patient treatment. *Proceedings of the Institution of Mechanical Engineers, Part H: Journal of Engineering in Medicine*, 236(11):1662–1674.

Chater, N. and Manning, C. D. (2006). Probabilistic models of language processing and acquisition. *Trends in cognitive sciences*, 10(7):335–344.

Chen, J., Shi, Y., Yi, C., Du, H., Kang, J., and Niyato, D. (2024a). Generative ai-driven human digital twin in iot-healthcare: A comprehensive survey. *arXiv preprint arXiv:2401.13699*.

Chen, J., Yi, C., Du, H., Niyato, D., Kang, J., Cai, J., and Shen, X. (2024b). A revolution of personalized healthcare: Enabling human digital twin with mobile aigc. *IEEE Network*.

Chen, J., Yi, C., Okegbile, S. D., Cai, J., Xuemin, and Shen (2024c). Networking Architecture and Key Supporting Technologies for Human Digital Twin in Personalized Healthcare: A Comprehensive Survey. *IEEE Communications Surveys & Tutorials*, 26(1):706–746.

Chen, X. (2021). Research on the Reform of Hybrid Teaching Mode Based on Digital Twin. In *2021 16th International Conference on Computer Science & Education (ICCSE)*, pages 672–675, Lancaster, United Kingdom. IEEE.

Cheng, Z. (2022). Human digital twin with applications. In *Proceedings of the 7th International Digital Human Modeling Symposium* -. University of Iowa Libraries Publishing.

Cheng, Z., Zientara, G. P., and Hoyt, R. W. (2019). Warfighter digital twin. *Defense TechConnect Innovation Summit & Showcase*, pages 8–10.

Chude-Okonkwo, U. K. (2021). Conceptual Molecular Communication Solution for Developing Digital Twin to Enable Precision Medicine Implementation. In *2021 15th International Conference on Signal Processing and Communication Systems (ICSPCS)*, pages 1–10, Sydney, Australia. IEEE.

Constantinescu, C., Rus, R., Rusu, C.-A., and Popescu, D. (2019). Digital Twins of Exoskeleton-Centered Workplaces: Challenges and Development Methodology. *Procedia Manufacturing*, 39:58–65.

Corbier, F. (1989). *Modélisation et émulation de la partie opérative pour la recette en plateforme d'équipements automatisés*. PhD thesis, Nancy 1.

Daily, L. Z., Lovett, M. C., and Reder, L. M. (2001). Modeling individual differences in working memory performance: A source activation account. *Cognitive Science*, 25(3):315–353.

Davis, F. D. et al. (1989). Technology acceptance model: Tam. *Al-Suqri, MN, Al-Aufi, AS: Information seeking behavior and technology adoption*, 205(219):5.

De Benedictis, A., Mazzocca, N., Somma, A., and Strigaro, C. (2023). Digital Twins in Healthcare: An Architectural Proposal and Its Application in a Social Distancing Case Study. *IEEE Journal of Biomedical and Health Informatics*, 27(10):5143–5154.

De Giglio, V., Evangelista, A., Uva, A. E., and Manghisi, V. M. (2026). Evaluating cognitive workload in industrial tasks via immersive vr and human digital twin frameworks: Development of application and design of experiment. *Procedia Computer Science*, 277:2828–2837.

Dell'Oca, S., Montini, E., Cutrona, V., Matteri, D., Landolfi, G., and Canetta, L. (2026). Beyond prediction accuracy in human-robot collaboration: Designing intention-aware cobot behavior (preprint).

Dimov, C., Khader, P. H., Marewski, J. N., and Pachur, T. (2020). How to model the neurocognitive dynamics of decision making: A methodological primer with act-r. *Behavior research methods*, 52(2):857–880.

Doerig, A., Sommers, R. P., Seeliger, K., Richards, B., Ismael, J., Lindsay, G. W., Kording, K. P., Konkle, T., Van Gerven, M. A., Kriegeskorte, N., et al. (2023). The neuroconnectionist research programme. *Nature Reviews Neuroscience*, 24(7):431–450.

Donatiello, L. and Marfia, G. (2022). Proposing the RecursiVerse Overlay Application for the MetaVerse. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 844–845, Christchurch, New Zealand. IEEE.

Doste, R., Lozano, M., Jimenez-Perez, G., Mont, L., Berruezo, A., Penela, D., Camara, O., and Sebastian, R. (2022). Training machine learning models with synthetic data improves the prediction of ventricular origin in outflow tract ventricular arrhythmias. *Frontiers in Physiology*, 13:909372.

Dragan, A. and Srinivasa, S. (2014). Integrating human observer inferences into robot motion planning. *Autonomous Robots*, 37:351–368.

Driess, D., Xia, F., Sajjadi, M. S., Lynch, C., Chowdhery, A., Ichter, B., Wahid, A., Tompson, J., Vuong, Q., Yu, T., et al. (2023). Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378*.

Du, Y., Zhang, B., Zheng, C., and Tang, Y. (2026). Towards seamless and safe human-robot collaboration in industry 5.0: advances in human behaviour prediction. *International Journal of Production Research*, pages 1–38.

Eliasmith, C., Stewart, T. C., Choo, X., Bekolay, T., DeWolf, T., Tang, Y., and Rasmussen, D. (2012). A large-scale model of the functioning brain. *science*, 338(6111):1202–1205.

Endo, Y., Maruyama, T., and Tada, M. (2023). DhaibaWorks: A Software Platform for Human-Centered Cyber-Physical Systems. *International Journal of Automation Technology*, 17(3):292–304.

European Commission and Directorate-General for Digital Services (2017). *New European interoperability framework – Promoting seamless services and data flows for European public administrations*. Publications Office.

Feng, B., Wang, Z., Yuan, L., Zhou, Q., Chen, Y., and Bi, Y. (2025). Towards safe motion planning for industrial human-robot interaction: A co-evolution approach based on human digital twin and mixed reality. *Robotics and Computer-Integrated Manufacturing*, 95:103012.

Ferrari, D., Rizzi, C., and Vitali, A. (2025). Human digital twin for realistic fall simulation using the behavioral foundation model meta motivo. In *International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, volume 89213, page V02BT02A037. American Society of Mechanical Engineers.

Fintz, M., Osadchy, M., and Hertz, U. (2022). Using deep learning to predict human decisions and using cognitive models to explain deep learning models. *Scientific reports*, 12(1):4736.

Flacco, F., Kroeger, T., De Luca, A., and Khatib, O. (2015). A depth space approach for evaluating distance to objects: with application to human-robot collision avoidance. *Journal of Intelligent & Robotic Systems*, 80:7–22.

Fodor, J. A. (1975). *The language of thought*, volume 5. Harvard university press.

Fratta, M., Floris, A., Porcu, S., and Atzori, L. (2026). The human digital twin for service management: Architecture and user modelling. *Computer Communications*, page 108526.

Furini, M., Gaggi, O., Mirri, S., Montangero, M., Pelle, E., Poggi, F., and Prandi, C. (2022). Digital twins and artificial intelligence: As pillars of personalized learning models. *Communications of the ACM*, 65(4):98–104.

Gaffinet, B., Ali, J. A. H., Naudet, Y., and Panetto, H. (2025a). Human digital twins: A systematic literature review and concept disambiguation for industry 5.0. *Computers in Industry*, 166:104230.

Gaffinet, B., Ali, J. A. H., Panetto, H., and Naudet, Y. (2023). Human-centric digital twins: Advancing safety and ergonomics in human-robot collaboration. In *International conference on innovative intelligent industrial production and logistics*, pages 380–397. Springer.

Gaffinet, B., Ali, J. A. H., Panetto, H., and Naudet, Y. (2025b). Towards cognitive interoperability with cognitive human digital twins. In *International conference on innovative intelligent industrial production and logistics*. Springer.

Gaffinet, B., Naudet, Y., and Panetto, H. (2025c). A systemic human digital twin model for human-centric systems. In *11th IFAC Conference on Manufacturing Modelling, Management and Control (MIM2025)*.

Gnecco, V. M., Vittori, F., and Pisello, A. L. (2023). Digital twins for decoding human-building interaction in multi-domain test-rooms for environmental comfort and energy saving via graph representation. *Energy and Buildings*, 279:112652.

Goldberg, L. R. (1993). The structure of phenotypic personality traits. *American psychologist*, 48(1):26.

Graessler, I. and Poehler, A. (2017). Integration of a digital twin as human representation in a scheduling procedure of a cyber-physical production system. In *2017 IEEE International Conference on Industrial Engineering and Engineering Management (IEEM)*, pages 289–293, Singapore. IEEE.

Graessler, I. and Poehler, A. (2018). Intelligent control of an assembly station by integration of a digital twin for employees into the decentralized control system. *Procedia Manufacturing*, 24:185–189.

Greco, A., Caterino, M., Fera, M., and Gerbino, S. (2020). Digital Twin for Monitoring Ergonomics during Manufacturing Production. *Applied Sciences*, 10(21):7758.

Grieves, M. (2002). Digital twin of physical systems: opportunities and challenges. In *Proceedings of the ASME 2002 International Mechanical Engineering Congress and Exposition, New Orleans, LA, USA*, pages 17–22.

Grieves, M. and Vickers, J. (2017). Digital Twin: Mitigating Unpredictable, Undesirable Emergent Behavior in Complex Systems. In Kahlen, F.-J., Flumerfelt, S., and Alves, A., editors, *Transdisciplinary Perspectives on Complex Systems*, pages 85–113. Springer International Publishing, Cham.

Gunzelmann, G., Gross, J. B., Gluck, K. A., and Dinges, D. F. (2009). Sleep deprivation and sustained attention performance: Integrating mathematical and cognitive modeling. *Cognitive science*, 33(5):880–910.

Haddadin, S., Albu-Schaffer, A., De Luca, A., and Hirzinger, G. (2008a). Collision detection and reaction: A contribution to safe physical human-robot interaction. In *2008 IEEE/RSJ international conference on intelligent robots and systems*, pages 3356–3363. IEEE.

Haddadin, S., Albu-Schäffer, A., and Hirzinger, G. (2009). Requirements for safe robots: Measurements, analysis and new insights. *The International Journal of Robotics Research*, 28(11-12):1507–1527.

Haddadin, S., Albu-Schäffer, A., Hirzinger, G., et al. (2007). Safety evaluation of physical human-robot interaction via crash-testing. In *Robotics: Science and systems*, volume 3, pages 217–224.

Haddadin, S., Albu-Schäffer, A. O., De Luca, A., and Hirzinger, G. (2008b). Evaluation of collision detection and reaction for a human-friendly robot on biological tissues. *Proceedings IARP 2008*.

Haddadin, S., Haddadin, S., Khoury, A., Rokahr, T., Parusel, S., Burgkart, R., Bicchi, A., and Albu-Schäffer, A. (2012). On making robots understand safety: Embedding injury knowledge into control. *The International Journal of Robotics Research*, 31(13):1578–1602.

Hafez, W. (2021). Human Digital Twin—Enabling Human-Agents Collaboration. In *2021 4th International Conference on Intelligent Robotics and Control Engineering (IRCE)*, pages 40–45, Lanzhou, China. IEEE.

Hancock, P. A., Billings, D. R., Schaefer, K. E., Chen, J. Y., De Visser, E. J., and Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5):517–527.

Hawkins, K. P., Vo, N., Bansal, S., and Bobick, A. F. (2013). Probabilistic human action prediction and wait-sensitive planning for responsive human-robot collaboration. In *2013 13th IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, pages 499–506. IEEE.

He, Q., Li, L., Li, D., Peng, T., Zhang, X., Cai, Y., Zhang, X., and Tang, R. (2024). From digital human modeling to human digital twin: Framework and perspectives in human factors. *Chinese Journal of Mechanical Engineering*, 37(1):9.

Hernandez-Boussard, T., Macklin, P., Greenspan, E. J., Gryshuk, A. L., Stahlberg, E., Syeda-Mahmood, T., and Shmulevich, I. (2021). Digital twins for predictive oncology will be a paradigm shift for precision cancer care. *Nature Medicine*, 27(12):2065–2066.

Hernigou, P., Olejnik, R., Safar, A., Martinov, S., Hernigou, J., and Ferre, B. (2021). Digital twins, artificial intelligence, and machine learning technology to identify a real personalized motion axis of the tibiotalar joint for robotics in total ankle arthroplasty. *International Orthopaedics*, 45(9):2209–2217.

Hu, Z., Lou, S., Xing, Y., Wang, X., Cao, D., and Lv, C. (2022). Review and Perspectives on Driver Digital Twin and Its Enabling Technologies for Intelligent Vehicles. *IEEE Transactions on Intelligent Vehicles*, 7(3):417–440.

Inamura, T. (2023). Digital Twin of Experience for Human–Robot Collaboration Through Virtual Reality. *International Journal of Automation Technology*, 17(3):284–291.

Inamura, T., Eitoku, S., Toshima, I., Shimizu, S., Fukayama, A., Ozawa, S., and Nakamura, T. (2022). Effect of repetitive motion intervention on self-avatar on the sense of self-individuality. In *Proceedings of the 10th International Conference on Human-Agent Interaction*, pages 167–175, Christchurch New Zealand. ACM.

Iqbal, J. D., Krauthammer, M., and Biller-Andorno, N. (2022). The Use and Ethics of Digital Twins in Medicine. *Journal of Law, Medicine & Ethics*, 50(3):583–596.

ISO-23247 (2020). Automation systems and integration—digital twin framework for manufacturing—part 1: overview and general principles. Standard, International Organization for Standardization, Geneva, CH.

ISO13855, E. (2010). Safety of machinery: Positioning of safeguards with respect to the approach speeds of parts of the human body. *International Organization for Standardization. Geneva.*

ISO/TS 15066, E. (2016). Iso/ts 15066 robots and robotic devices - collaborative robots. *International Organization for Standardization. Geneva.*

Iung, B., Morel, D., and Lhoste, P. (2000). Interoperable component-based emulation for validation of distributed control. *IFAC Proceedings Volumes*, 33(20):159–165.

Ivanov, D. (2023). The industry 5.0 framework: viability-based integration of the resilience, sustainability, and human-centricity perspectives. *International Journal of Production Research*, 61(5):1683–1695.

Iwaki, T., Satoshi, K., Hajime, N., Takao, K., Keiichi, H., and Shiro, O. (2020). Challenges facing human digital twin computing and its future prospects. *NTT Technical Review*, 18(9):19–24.

Jaleel, S. (2022). The Metaverse – A universe of human digital clones. In *2022 IEEE International Conference on Metrology for Extended Reality, Artificial Intelligence and Neural Engineering (MetroXRAINE)*, pages 81–86, Rome, Italy. IEEE.

Jensen, J. and Deng, J. (2023). Digital Twins for Radiation Oncology. In *Companion Proceedings of the ACM Web Conference 2023*, pages 989–993, Austin TX USA. ACM.

Ji, T., Polzer, J., and Xu, X. (2023). Cognitive digital twin framework for smart manufacturing. In *2023 IEEE 19th International Conference on Automation Science and Engineering (CASE)*, pages 1–6. IEEE.

Joseph, A. J., Kruger, K., and Basson, A. H. (2021). An Aggregated Digital Twin Solution for Human-Robot Collaboration in Industry 4.0 Environments. In Borangiu, T., Trentesaux, D., Leitão, P., Cardin, O., and Lamouri, S., editors, *Service Oriented, Holonic and Multi-Agent Manufacturing Systems for Industry of the Future*, volume 952, pages 135–147. Springer International Publishing, Cham.

Kazerouni, A., Aghdam, E. K., Heidari, M., Azad, R., Fayyaz, M., Hacihaliloglu, I., and Merhof, D. (2023). Diffusion models in medical imaging: A comprehensive survey. *Medical Image Analysis*, 88:102846.

Kider, J. T. (2025). *Human-Centered Digital Twins*, pages 1–37. Springer Nature Singapore, Singapore.

Kieras, D. E. and Meyer, D. E. (1997). An overview of the epic architecture for cognition and performance with application to human-computer interaction. *Human–Computer Interaction*, 12(4):391–438.

Kim, J. (2007). Physicalism, or something near enough. *Princeton University Press*.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kobayashi, T., Fukae, K., Imai, T., and Arai, K. (2022). Digital Twin Agent for Super-Aged Society. In *2022 IEEE International Conference on Consumer Electronics (ICCE)*, pages 1–6, Las Vegas, NV, USA. IEEE.

Kopponen, A., Hahto, A., Kettunen, P., Mikkonen, T., Makitalo, N., Nurmi, J., and Rossi, M. (2022). Empowering Citizens With Digital Twins: A Blueprint. *IEEE Internet Computing*, 26(5):7–16.

Kotseruba, I. and Tsotsos, J. K. (2020). 40 years of cognitive architectures: core cognitive abilities and practical applications. *Artificial Intelligence Review*, 53(1):17–94.

Kritzinger, W., Karner, M., Traar, G., Henjes, J., and Sihn, W. (2018). Digital twin in manufacturing: A categorical literature review and classification. *IFAC-PapersOnLine*, 51(11):1016–1022. 16th IFAC Symposium on Information Control Problems in Manufacturing INCOM 2018.

Krüger, J., Lien, T. K., and Verl, A. (2009). Cooperation of human and machines in assembly lines. *CIRP annals*, 58(2):628–646.

Kulic, D. and Croft, E. A. (2007). Affective state estimation for human–robot interaction. *IEEE transactions on robotics*, 23(5):991–1000.

L Griffiths, T., Kemp, C., and B Tenenbaum, J. (2008). *Bayesian models of cognition*. Carnegie Mellon University.

Laffranchi, M., Tsagarakis, N. G., and Caldwell, D. G. (2009). Safe human robot interaction via energy regulation control. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 35–41. IEEE.

Laird, J. E. (2019). *The Soar cognitive architecture*. MIT press.

Laird, J. E., Kinkade, K. R., Mohan, S., and Xu, J. Z. (2012). Cognitive robotics using the soar cognitive architecture. In *CogRob@ AAAI*.

Laird, J. E., Lebiere, C., and Rosenbloom, P. S. (2017). A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics. *Ai Magazine*, 38(4):13–26.

Lakens, D. (2022). Sample size justification. *Collabra: psychology*, 8(1):33267.

Langås, E. F., Zafar, M. H., and Sanfilippo, F. (2026). Exploring the synergy of human-robot teaming, digital twins, and machine learning in industry 5.0: a step towards sustainable manufacturing. *Journal of Intelligent Manufacturing*, 37(3):999–1022.

Lauer-Schmaltz, M. W., Cash, P., Hansen, J. P., and Maier, A. (2022). Designing Human Digital Twins for Behaviour-Changing Therapy and Rehabilitation: A Systematic Review. *Proceedings of the Design Society*, 2:1303–1312.

Lauer-Schmaltz, M. W., Cash, P., Hansen, J. P., and Maier, A. (2024). Towards the human digital twin: Definition and design–a survey. *arXiv preprint arXiv:2402.07922*.

Le Moigne, J.-L. (1990). *La Modelisation des Systemes Complexes*.

Lebiere, C. (1999). The dynamics of cognition: An act-r model of cognitive arithmetic. *Kognitionswissenschaft*, 8(1):5–19.

Lee, H., Kim, S. D., and Amin, M. A. U. A. (2022). Control framework for collaborative robot using imitation learning-based teleoperation from human digital twin to robot digital twin. *Mechatronics*, 85:102833.

Lezoche, M. and Panetto, H. (2020). Cyber-physical systems, a new formal paradigm to model redundancy and resiliency. *Enterprise Information Systems*, 14(8):1150–1171.

Liao, X., Zhao, X., Wang, Z., Zhao, Z., Han, K., Gupta, R., Barth, M. J., and Wu, G. (2023). Driver Digital Twin for Online Prediction of Personalized Lane-Change Behavior. *IEEE Internet of Things Journal*, 10(15):13235–13246.

Lin, Y., Chen, L., Ali, A., Nugent, C., Cleland, I., Li, R., Ding, J., and Ning, H. (2024). Human digital twin: A survey. *Journal of Cloud Computing*, 13(1):131.

Lin, Y., Chen, L., Ali, A., Nugent, C., Ian, C., Li, R., Gao, D., Wang, H., Wang, Y., and Ning, H. (2022). Human digital twin: A survey. *arXiv preprint arXiv:2212.05937*.

Linardatos, P., Papastefanopoulos, V., and Kotsiantis, S. (2020). Explainable ai: A review of machine learning interpretability methods. *Entropy*, 23(1):18.

Liu, H., Wang, Y., Ji, W., and Wang, L. (2018). A context-aware safety system for human-robot collaboration. *Procedia Manufacturing*, 17:238–245.

Liu, X., Li, G., Xiang, F., Tao, B., and Jiang, G. (2025). A digital twin-driven method for improving human comfort in human–robot collaboration. *The International Journal of Advanced Manufacturing Technology*, 137(1):339–359.

Liu, X., Zheng, L., Wang, Y., Yang, W., Jiang, Z., Wang, B., Tao, F., and Li, Y. (2022). Human-centric collaborative assembly system for large-scale space deployable mechanism driven by Digital Twins and wearable AR devices. *Journal of Manufacturing Systems*, 65:720–742.

Löcklin, A., Jung, T., Jazdi, N., Ruppert, T., and Weyrich, M. (2021). Architecture of a Human-Digital Twin as Common Interface for Operator 4.0 Applications. *Procedia CIRP*, 104:458–463.

Lonsdale, H., Gray, G. M., Ahumada, L. M., Yates, H. M., Varughese, A., and Rehman, M. A. (2022). The Perioperative Human Digital Twin. *Anesthesia & Analgesia*, 134(4):885–892.

Lorenzini, M., Lagomarsino, M., Fortini, L., Gholami, S., and Ajoudani, A. (2023). Ergonomic human-robot collaboration in industry: A review. *Frontiers in Robotics and AI*, 9:813907.

Lovett, M. C., Daily, L. Z., and Reder, L. M. (2000). A source activation theory of working memory: Cross-task prediction of performance in act-r. *Cognitive Systems Research*, 1(2):99–118.

Lovett, M. C., Reder, L. M., and Lebiere, C. (1997). Modeling individual differences in a digit working memory task. In *Proceedings of the Nineteenth Annual Conference of the Cognitive Science Society*, volume 19, page 460.

Lovett, M. C., Reder, L. M., and Lebiere, C. (1999). Modeling working memory in a unified architecture: An act-r perspective. *Models of working memory: Mechanisms of active maintenance and executive control*, 506:135–182.

Lu, Y., Zheng, H., Chand, S., Xia, W., Liu, Z., Xu, X., Wang, L., Qin, Z., and Bao, J. (2022). Outlook on human-centric manufacturing towards industry 5.0. *Journal of Manufacturing Systems*, 62:612–627.

Marewski, J. N. and Mehlhorn, K. (2011). Using the act-r architecture to specify 39 quantitative process models of decision making. *Judgment and Decision making*, 6(6):439–519.

Martynov, V., Filosova, E., and Egorova, Y. (2022). Information Architecture to Support Engineering Education in the Era of Industry 4.0. In *2022 VI International Conference on Information Technologies in Engineering Education (Inforino)*, pages 1–5, Moscow, Russian Federation. IEEE.

Maruyama, T., Ueshiba, T., Tada, M., Toda, H., Endo, Y., Domae, Y., Nakabo, Y., Mori, T., and Suita, K. (2021). Digital twin-driven human robot collaboration using a digital human. *Sensors*, 21(24):8266.

Meguenani, A., Padois, V., Da Silva, J., Hoarau, A., and Bidaud, P. (2016). Energy based control for safe human-robot physical interaction. In *International Symposium on Experimental Robotics*, pages 809–818. Springer.

Mertens, A., Putz, S., Brauner, P., Brillowski, F., Buczak, N., Dammers, H., Van Dyck, M., Kong, I., Konigs, P., Kordtomeikel, F., Rodemann, N., Schaar, A. K., Steuer-Dankert, L., Wlecke, S., Gries, T., Leicht-Scholten, C., Nagel, S. K., Piller, F. T., Schuh, G., Ziefle, M., and Nitsch, V. (2021). Human Digital Shadow: Data-based Modeling of Users and Usage in the Internet of Production. In *2021 14th International Conference on Human System Interaction (HSI)*, pages 1–8, Gdańsk, Poland. IEEE.

Miller, M. E. and Spatz, E. (2022a). A unified view of a human digital twin. *Human-Intelligent Systems Integration*, 4(1-2):23–33.

Miller, M. E. and Spatz, E. (2022b). A unified view of a human digital twin. *Human-Intelligent Systems Integration*, 4(1):23–33.

Mohammed, A., Schmidt, B., and Wang, L. (2017). Active collision avoidance for human–robot collaboration driven by vision sensors. *International Journal of Computer Integrated Manufacturing*, 30(9):970–980.

Montini, E., Bettoni, A., Ciavotta, M., Carpanzano, E., and Pedrazzoli, P. (2021a). A metamodel for modular composition of tailored human digital twins in production. *Procedia CIRP*, 104:689–695.

Montini, E., Bonomi, N., Daniele, F., Bettoni, A., Pedrazzoli, P., Carpanzano, E., Rocco, P., et al. (2021b). The human-digital twin in the manufacturing industry: Current perspectives and a glimpse of future. *Trusted artificial intelligence in manufacturing: A review of the emerging wave of ethical and human centric AI technologies for smart production*, pages 132–147.

Montini, E., Cutrona, V., Bonomi, N., Landolfi, G., Bettoni, A., Rocco, P., and Carpanzano, E. (2022). An IIoT Platform For Human-Aware Factory Digital Twins. *Procedia CIRP*, 107:661–667.

Morris, M. B., Swan, G., Veksler, B. Z., Tuttle, J., and Stephenson, A. (2025). Integrating individual differences metrics into cognitive models. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, page 10711813251369464. SAGE Publications Sage CA: Los Angeles, CA.

Mourtzis, D., Angelopoulos, J., Panopoulos, N., and Kardamakis, D. (2021). A Smart IoT Platform for Oncology Patient Diagnosis based on AI: Towards the Human Digital Twin. *Procedia CIRP*, 104:1686–1691.

Naudet, Y., Ali, J. A. H., Gaffinet, B., and Panetto, H. (2024). Cognition in digital twins for cyber-physical systems and humans: Where and why? In *International Conference on Innovative Intelligent Industrial Production and Logistics*, pages 399–412. Springer.

Naudet, Y., Antoniou, A., Lykourentzou, I., Tobias, E., Rompa, J., and Lepouras, G. (2015). Museum personalization based on gaming and cognitive styles: The blue experiment. *Int. J. Virtual Communities Soc. Netw.*, 7:1–30.

Naudet, Y., Baudet, A., and Risse, M. (2021a). Human Digital Twin in Industry 4.0: Concept and Preliminary Model:. In *Proceedings of the 2nd International Conference on Innovative Intelligent Industrial Production and Logistics*, pages 137–144.

Naudet, Y., Baudet, A., and Risse, M. (2021b). Human digital twin in industry 4.0: Concept and preliminary model. In *IN4PL*, pages 137–144.

Naudet, Y., Zehnder, E., Gaffinet, B., Al Haj Ali, J., and Panetto, H. (2025). A human ontology with cognition and ability models for human and cognitive digital twins. In *International Conference on Innovative Intelligent Industrial Production and Logistics*, pages 52–71. Springer.

Neisser, U. (1967). *Cognitive psychology: Classic edition*. Psychology press.

Newell, A. (1990). *Unified theories of cognition*. Harvard University Press.

Newell, A. and Simon, H. A. (1975). Computer science as empirical inquiry: Symbols and search. In *ACM Turing award lectures*.

Nguyen, T. N. (2021). Cybonto: Towards Human Cognitive Digital Twins for Cybersecurity. *arXiv preprint arXiv:2108.00551*.

Nguyen, T. N. (2022). Toward Human Digital Twins for Cybersecurity Simulations on the Metaverse: Ontological and Network Science Approach. *JMIRx Med*, 3(2):e33502.

Nikolaidis, S. and Shah, J. (2012). Human-robot interactive planning using cross-training: A human team training approach. In *Infotech@ Aerospace 2012*, page 2536. AIAA.

Ogunseiju, O. R., Olayiwola, J., Akanmu, A. A., and Nnaji, C. (2021). Digital twin-driven framework for improving self-management of ergonomic risks. *Smart and Sustainable Built Environment*, 10(3):403–419.

Okegbile, S., Cai, J., Wu, J., Chen, J., and Yi, C. (2023a). A prediction-enhanced physical-to-virtual twin connectivity framework for human digital twin. *Authorea Preprints*.

Okegbile, S. D. and Cai, J. (2022). Edge-assisted human-to-virtual twin connectivity scheme for human digital twin frameworks. In *2022 IEEE 95th Vehicular Technology Conference: (VTC2022-Spring)*, pages 1–6, Helsinki, Finland. IEEE.

Okegbile, S. D., Cai, J., Chen, J., and Yi, C. (2024). A reputation-enhanced shard-based byzantine fault-tolerant scheme for secure data sharing in zero trust human digital twin systems. *IEEE Internet of Things Journal*.

Okegbile, S. D., Cai, J., Niyato, D., and Yi, C. (2023b). Human Digital Twin for Personalized Healthcare: Vision, Architecture and Future Directions. *IEEE Network*, 37(2):262–269.

Okegbile, S. D., Cai, J., Zheng, H., Chen, J., and Yi, C. (2023c). Differentially private federated multi-task learning framework for enhancing human-to-virtual connectivity in human digital twin. *IEEE Journal on Selected Areas in Communications*.

Olfati-Saber, R. and Murray, R. M. (2004). Consensus problems in networks of agents with switching topology and time-delays. *IEEE Transactions on automatic control*, 49(9):1520–1533.

Omodei, M. M. and Wearing, A. J. (1995). The fire chief microworld generating program: An illustration of computer-simulated microworlds as an experimental paradigm for studying complex decision-making behavior. *Behavior Research Methods, Instruments, & Computers*, 27(3):303–316.

Oztemel, E. and Gursev, S. (2020). Literature review of industry 4.0 and related technologies. *Journal of intelligent manufacturing*, 31(1):127–182.

O'Reilly, R. C., Hazy, T. E., and Herd, S. A. (2012). *The leabra cognitive architecture: how to play 20 principles with nature and win.* Oxford University Press, Oxford, UK.

Pacaux-Lemoine, M.-P., Trentesaux, D., Rey, G. Z., and Millot, P. (2017). Designing intelligent manufacturing systems through human-machine cooperation principles: a human-centered approach. *Computers & Industrial Engineering*, 111:581–595.

Pandey, A. K. and Alami, R. (2010). Mightability maps: A perceptual level decisional framework for co-operative and competitive human-robot interaction. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5842–5848. IEEE.

Panetto, H. (1991). *Une contribution au Génie Automatique: le prototypage des machines et systèmes automatisés de production.(Contribution to Automation Engineering: Prototyping of Production Systems and Machines).* PhD thesis, Henri Poincaré University, Nancy, France.

Park, J.-J., Lee, Y.-J., Song, J.-B., and Kim, H.-S. (2008). Safe joint mechanism based on nonlinear stiffness for safe human-robot collision. In *2008 IEEE International Conference on Robotics and Automation*, pages 2177–2182. IEEE.

Park, J. S., O'Brien, J., Cai, C. J., Morris, M. R., Liang, P., and Bernstein, M. S. (2023). Generative agents: Interactive simulacra of human behavior. In *Proceedings of the 36th annual acm symposium on user interface software and technology*, pages 1–22.

Passera, K., Manini, S., Antiga, L., and Remuzzi, A. (2013). Patient-specific model of arterial circulation for surgical planning of vascular access. *The journal of vascular access*, 14(2):180–192.

Pellegrinelli, S., Admoni, H., Javdani, S., and Srinivasa, S. (2016). Human-robot shared workspace collaboration via hindsight optimization. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 831–838. IEEE.

Pesapane, F., Rotili, A., Penco, S., Nicosia, L., and Cassano, E. (2022). Digital Twins in Radiology. *Journal of Clinical Medicine*, 11(21):6553.

Prokscha, A., Sheikh, F., Lessy, D., Ali, N. A., and Kaiser, T. (2022a). Multidisciplinary Data Fusion for THz Ray- Tracing. In *2022 16th European Conference on Antennas and Propagation (EuCAP)*, pages 1–5, Madrid, Spain. IEEE.

Prokscha, A., Sheikh, F., Lessy, D., and Kaiser, T. (2022b). Portable Device-Centric Human Postures for Modeling Reliable THz Channels. In *2022 Fifth International Workshop on Mobile Terahertz Systems (IWMTS)*, pages 1–5, Duisburg, Germany. IEEE.

Qi, Q., Tao, F., Hu, T., Anwer, N., Liu, A., Wei, Y., Wang, L., and Nee, A. (2021). Enabling technologies and tools for digital twin. *Journal of Manufacturing Systems*, 58:3–21.

Qu, Y., Zhao, N., and Zhang, H. (2024). Digital twin technology of human–machine integration in cross-belt sorting system. *Chinese Journal of Mechanical Engineering*, 37(1):33.

Ramasubramanian, A. K., Mathew, R., Kelly, M., Hargaden, V., and Papakostas, N. (2022). Digital Twin for Human–Robot Collaboration in Manufacturing: Review and Outlook. *Applied Sciences*, 12(10):4811.

Ramos, D., Goncalves, G., Faria, R., and Sanches, M. P. (2021). Building a Drone Operator Digital Twin using a Brain-Computer Interface for Emotion Recognition. In *2021 20th International Conference on Advanced Robotics (ICAR)*, pages 824–829, Ljubljana, Slovenia. IEEE.

Ras, G., Xie, N., Van Gerven, M., and Doran, D. (2022). Explainable deep learning: A field guide for the uninitiated. *Journal of Artificial Intelligence Research*, 73:329–396.

Reed, M. P., Faraway, J., Chaffin, D. B., and Martin, B. J. (2006). The humosim ergonomics framework: A new approach to digital human simulation for ergonomic analysis. Technical report, SAE Technical Paper.

Regazzoni, D. and Rizzi, C. (2019). Virtualization of the human in the digital factory. *Systems engineering in the fourth industrial revolution*, pages 161–189.

Rehling, J., Lovett, M., Lebiere, C., Reder, L., and Demiral, B. (2004). Modeling complex tasks: An individual difference approach. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 26.

Ritter, F. E., Tehranchi, F., and Oury, J. D. (2019). Act-r: A cognitive architecture for modeling cognition. *Wiley Interdisciplinary Reviews: Cognitive Science*, 10(3):e1488.

Rivera, L. F., Jiménez, M., Angara, P., Villegas, N. M., Tamura, G., and Müller, H. A. (2019). Towards Continuous Monitoring in Personalized Healthcare through Digital Twins. In *CASCON '19: Proceedings of the 29th Annual International Conference on Computer Science and Software Engineering*.

Romero, D., Bernus, P., Noran, O., Stahre, J., and Fast-Berglund, Å. (2016). The operator 4.0: Human cyber-physical systems & adaptive automation towards human-automation symbiosis work systems. In *IFIP international conference on advances in production management systems*, pages 677–686. Springer.

Romero, O. J., Zimmerman, J., Steinfeld, A., and Tomasic, A. (2023). Synergistic integration of large language models and cognitive architectures for robust ai: An exploratory analysis. In *Proceedings of the AAAI Symposium Series*, volume 2, pages 396–405.

Rosenblatt, F. (1958). The perceptron: a probabilistic model for information storage and organization in the brain. *Psychological review*, 65(6):386.

Rumelhart, D. E., Hinton, G. E., and Williams, R. J. (1986). Learning representations by back-propagating errors. *nature*, 323(6088):533–536.

Sahal, R., Alsamhi, S. H., and Brown, K. N. (2022). Personal Digital Twin: A Close Look into the Present and a Step towards the Future of Personalised Healthcare Industry. *Sensors*, 22(15):5918.

Salvucci, D. D. (2000). A model of eye movements and visual attention. In *Proceedings of the international conference on cognitive modeling*, number S 252, page 259.

Sauro, J. and Dumas, J. S. (2009). Comparison of three one-question, post-task usability questionnaires. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pages 1599–1608.

Semeraro, C., Lezoche, M., Panetto, H., and Dassisti, M. (2021). Digital twin paradigm: A systematic literature review. *Computers in Industry*, 130:103469.

Shangguan, D., Chen, L., Su, C., Ding, J., and Liu, C. (2022). A Triple Human-Digital Twin Architecture for Cyber-Physical Systems. *Computer Modeling in Engineering & Sciences*, 131(3):1557–1578.

Sharotry, A., Jimenez, J. A., Mediavilla, F. A. M., Wierschem, D., Koldenhoven, R. M., and Valles, D. (2022). Manufacturing Operator Ergonomics: A Conceptual Digital Twin Approach to Detect Biomechanical Fatigue. *IEEE Access*, 10:12774–12791.

Shengli, W. (2021a). Is Human Digital Twin possible? *Computer Methods and Programs in Biomedicine Update*, 1:100014.

Shengli, W. (2021b). Is human digital twin possible? *Computer Methods and Programs in Biomedicine Update*, 1:100014.

Sinisi, S., Alimguzhin, V., Mancini, T., Tronci, E., Mari, F., and Leeners, B. (2020). Optimal Personalised Treatment Computation through In Silico Clinical Trials on Patient Digital Twins*. *Fundamenta Informaticae*, 174(3-4):283–310.

Sparrow, D., Kruger, K., and Basson, A. (2019). Human Digital Twin for integrating human workers in Industry 4.0. *COMA'19*, page 259.

Sparrow, D. E., Kruger, K., and Basson, A. H. (2022). An architecture to facilitate the integration of human workers in Industry 4.0 environments. *International Journal of Production Research*, 60(15):4778–4796.

Stacchio, L., Perlino, M., Vagnoni, U., Sasso, F., Scorolli, C., and Marfia, G. (2022). Who will Trust my Digital Twin? Maybe a Clerk in a Brick and Mortar Fashion Shop. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pages 814–815, Christchurch, New Zealand. IEEE.

Stewart, T., Choo, F.-X., and Eliasmith, C. (2012). Spaun: A perception-cognition-action model using spiking neurons. In *Proceedings of the Annual Meeting of the Cognitive Science Society*, volume 34.

Steyvers, M., Tenenbaum, J. B., Wagenmakers, E.-J., and Blum, B. (2003). Inferring causal networks from observations and interventions. *Cognitive science*, 27(3):453–489.

Stocco, A., Mitsopoulos, K., Yang, Y. C., Hake, H. S., Haile, T., Leonard, B., and Gluck, K. (2024). Fitting, evaluating, and comparing cognitive architecture models using likelihood: A primer with examples in act-r. *arXiv preprint arXiv:2410.18055*.

Stocco, A., Sibert, C., Steine-Hanson, Z., Koh, N., Laird, J. E., Lebiere, C. J., and Rosenbloom, P. (2021). Analysis of the human connectome data supports the notion of a "common model of cognition" for human and human-like intelligence across domains. *NeuroImage*, 235:118035.

Storn, R. and Price, K. (1997). Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. *Journal of global optimization*, 11(4):341–359.

Sun, R. (2006). The clarion cognitive architecture: Extending cognitive modeling to social simulation. *Cognition and multi-agent interaction*, pages 79–99.

Taatgen, N. A. (2002). A model of individual differences in skill acquisition in the kanfer–ackerman air traffic control task. *Cognitive Systems Research*, 3(1):103–112.

Taatgen, N. A., Lebiere, C., and Anderson, J. R. (2006). Modeling paradigms in act-r. *Cognition and multi-agent interaction: From cognitive modeling to social simulation*, pages 29–52.

Tan, J. T. C. and Arai, T. (2011). Triple stereo vision system for safety monitoring of human-robot collaboration in cellular manufacturing. In *2011 IEEE International Symposium on Assembly and Manufacturing (ISAM)*, pages 1–6. IEEE.

Tao, F., Cheng, J., Qi, Q., Zhang, M., Zhang, H., and Sui, F. (2018). Digital twin-driven product design, manufacturing and service with big data. *The International Journal of Advanced Manufacturing Technology*, 94(9-12):3563–3576.

Tardini, E., Zhang, X., Canahuate, G., Wentzel, A., Mohamed, A. S. R., Van Dijk, L., Fuller, C. D., and Marai, G. E. (2022). Optimal Treatment Selection in Sequential Systemic and Locoregional Therapy of Oropharyngeal Squamous Carcinomas: Deep Q-Learning With a Patient-Physician Digital Twin Dyad. *Journal of Medical Internet Research*, 24(4):e29455.

Taylor, N. C., Bekker, A., and Kruger, K. (2023). MARINER 4.0: INTEGRATING SEAFARERS INTO A MARITIME 4.0 ENVIRONMENT. *International Journal of Maritime Engineering*, 164(A4):373–384.

Thamotharan, P., Srinivasan, S., Kesavadev, J., Krishnan, G., Mohan, V., Seshadhri, S., Bekiroglu, K., and Toffanin, C. (2023). Human Digital Twin for Personalized Elderly Type 2 Diabetes Management. *Journal of Clinical Medicine*, 12(6):2094.

Tonietti, G., Schiavi, R., and Bicchi, A. (2005). Design and control of a variable stiffness actuator for safe and fast physical human/robot interaction. In *Proceedings of the 2005 IEEE international conference on robotics and automation*, pages 526–531. IEEE.

Trafton, J. G., Hiatt, L. M., Harrison, A. M., Tamborello, F. P., Khemlani, S. S., and Schultz, A. C. (2013). Act-r/e: An embodied cognitive architecture for human-robot interaction. *Journal of Human-Robot Interaction*, 2(1):30–55.

Trentesaux, D. and Millot, P. (2016). A human-centred design to break the myth of the "magic human" in intelligent manufacturing systems. In *Service orientation in holonic and multi-agent manufacturing*, pages 103–113. Springer.

Trobinger, M., Costinescu, A., Xing, H., Elsner, J., Hu, T., Naceri, A., Figueredo, L., Jensen, E., Burschka, D., and Haddadin, S. (2021). A Dual Doctor-Patient Twin Paradigm for Transparent Remote Examination, Diagnosis, and Rehabilitation. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 2933–2940, Prague, Czech Republic. IEEE.

Uhlenberg, L., Derungs, A., and Amft, O. (2023). Co-simulation of human digital twins and wearable inertial sensors to analyse gait event estimation. *Frontiers in Bioengineering and Biotechnology*, 11:1104000.

Unhelkar, V. V., Lasota, P. A., Tyroller, Q., Buhai, R.-D., Marceau, L., Deml, B., and Shah, J. A. (2018). Human-aware robotic assistant for collaborative assembly: Integrating human motion prediction with planning in time. *IEEE Robotics and Automation Letters*, 3(3):2394–2401.

van der Velde, M., Sense, F., Borst, J. P., van Maanen, L., and Van Rijn, H. (2022). Capturing dynamic performance in a cognitive model: Estimating act-r memory parameters with the linear ballistic accumulator. *Topics in Cognitive Science*, 14(4):889–903.

Vasic, M. and Billard, A. (2013). Safety issues in human-robot interactions. In *2013 ieee international conference on robotics and automation*, pages 197–204. IEEE.

Venkatesh, V., Morris, M. G., Davis, G. B., and Davis, F. D. (2003). User acceptance of information technology: Toward a unified view1. *MIS quarterly*, 27(3):425–478.

Von Bertalanffy, L. (1972). The history and status of general systems theory. *Academy of management journal*, 15(4):407–426.

Wang, B., Zhou, H., Li, X., Yang, G., Zheng, P., Song, C., Yuan, Y., Wuest, T., Yang, H., and Wang, L. (2024a). Human Digital Twin in the context of Industry 5.0. *Robotics and Computer-Integrated Manufacturing*, 85:102626.

Wang, B., Zhou, H., Li, X., Yang, G., Zheng, P., Song, C., Yuan, Y., Wuest, T., Yang, H., and Wang, L. (2024b). Human digital twin in the context of industry 5.0. *Robotics and Computer-Integrated Manufacturing*, 85:102626.

Wang, B., Zhou, H., Yang, G., Li, X., and Yang, H. (2022a). Human Digital Twin (HDT) Driven Human-Cyber-Physical Systems: Key Technologies and Applications. *Chinese Journal of Mechanical Engineering*, 35(1):11.

Wang, L., Gao, R., Váncza, J., Krüger, J., Wang, X. V., Makris, S., and Chryssolouris, G. (2019). Symbiotic human-robot collaborative assembly. *CIRP annals*, 68(2):701–726.

Wang, L., Schmidt, B., and Nee, A. Y. (2013). Vision-guided active collision avoidance for human-robot collaborations. *Manufacturing Letters*, 1(1):5–8.

Wang, S., Zhang, J., Wang, P., Law, J., Calinescu, R., and Mihaylova, L. (2024c). A deep learning-enhanced digital twin framework for improving safety and reliability in human–robot collaborative manufacturing. *Robotics and computer-integrated manufacturing*, 85:102608.

Wang, Z. (2021). Mobility Digital Twin with Connected Vehicles and Cloud Computing. *Preprint*. TechRxiv.

Wang, Z., Gupta, R., Han, K., Wang, H., Ganlath, A., Ammar, N., and Tiwari, P. (2022b). Mobility Digital Twin: Concept, Architecture, Case Study, and Future Challenges. *IEEE Internet of Things Journal*, 9(18):17452–17467.

Wang, Z., Liu, Y., and Hansen, J. H. (2023a). Enhancing driver visual guidance through mobility digital twin. *Towards Human-Vehicle Harmonization*, 3:95.

Wang, Z., Zheng, O., Li, L., Abdel-Aty, M., Cruz-Neira, C., and Islam, Z. (2023b). Towards Next Generation of Pedestrian and Connected Vehicle In-the-Loop Research: A Digital Twin Co-Simulation Framework. *IEEE Transactions on Intelligent Vehicles*, 8(4):2674–2683.

Wellman, M. P. (1993). A market-oriented programming environment and its application to distributed multicommodity flow problems. *Journal of artificial intelligence research*, 1:1–23.

Weyns, D. and Holvoet, T. (2004). A formal model for situated multi-agent systems. *Fundamenta Informaticae*, 63(2-3):125–158.

Wooldridge, M. (2009). *An introduction to multiagent systems*. John wiley & sons.

World Wide Web Consortium (2017). Semantic sensor network ontology. W3C Recommendation.

Wu, J., Wang, Z., Sarker, A., and Srivastava, M. (2023). Acuity: Creating Realistic Digital Twins Through Multi-resolution Pointcloud Processing and Audiovisual Sensor Fusion. In *Proceedings of the 8th ACM/IEEE Conference on Internet of Things Design and Implementation*, pages 79–92, San Antonio TX USA. ACM.

Xiang, H., Wu, K., Chen, J., Yi, C., Cai, J., Niyato, D., Xuemin, and Shen (2023). Edge Computing Empowered Tactile Internet for Human Digital Twin: Visions and Case Study. *Pre-print*. arXiv:2304.07454.

Xinzhou, W. and Kuzmichev, V. (2022). Digital twins for wetsuit design. In *Proceedings Of Higher Education Institutions. Textile Industry Technology*.

Yamada, Y., Hirasawa, Y., Huang, S., and Umetani, Y. (1996). Fail-safe human/robot contact in the safety space. In *Proceedings 5th IEEE International Workshop on Robot and Human Communication. RO-MAN'96 TSUKUBA*, pages 59–64. IEEE.

Yang, Y., Shi, Y., Yi, C., Cai, J., Kang, J., Niyato, D., and Shen, X. (2024). Dynamic human digital twin deployment at the edge for task execution: A two-timescale accuracy-aware online optimization. *IEEE Transactions on Mobile Computing*.

You, Y., Cai, B., Pham, D. T., Liu, Y., and Ji, Z. (2025). A human digital twin approach for fatigue-aware task planning in human-robot collaborative assembly. *Computers & Industrial Engineering*, 200:110774.

Yuille, A. and Kersten, D. (2006). Vision as bayesian inference: analysis by synthesis? *Trends in cognitive sciences*, 10(7):301–308.

Zehnder, E. and Naudet, Y. (2025). Human-centered cognition model for human digital twins. In *International Conference on Human-Computer Interaction*, pages 261–280. Springer.

Zhang, Z., Ji, Y., Tang, D., Chen, J., and Liu, C. (2023). Enabling Collaborative Assembly between Humans and Robots Using a Digital Twin System for Human-Robot Symbiosis. Preprint, SSRN.

Zhao, X. and Pan, J. (2018). Considering human behavior in motion planning for smooth human-robot collaboration in close proximity. In *2018 27th IEEE international symposium on robot and human interactive communication (RO-MAN)*, pages 985–990. IEEE.

Zheng, P., Li, S., Fan, J., Li, C., and Wang, L. (2023). A collaborative intelligence-based approach for handling human-robot collaboration uncertainties. *CIRP Annals*, 72(1):1–4.

Zheng, X., Lu, J., and Kiritsis, D. (2022). The emergence of cognitive digital twin: vision, challenges and opportunities. *International Journal of Production Research*, 60(24):7610–7632.

Zhong, R., Hu, B., Hong, Z., Zhang, Z., Lou, S., Song, X., Feng, Y., and Tan, J. (2025). Human-robot handover task intention recognition framework by fusing human digital twin and deep domain adaptation. *Journal of Engineering Design*, 36(11):2055–2071.

Zhou, W., Armstrong, T. J., Reed, M. P., Hoffman, S. G., and Wegner, D. M. (2009). Simulating complex automotive assembly tasks using the humosim framework. Technical report, SAE Technical Paper.

Zhu, E. and Yang, S. (2025). Towards human digital twin: Reviewing human modelling and simulation. *Journal of Industrial Information Integration*, page 100975.

Zibuschka, J., Ruff, C., Horch, A., and Roßnagel, H. (2020). A human digital twin as building block of open identity management for the internet of things. In *Open Identity Summit 2020*, pages 133–142. Gesellschaft für Informatik e.V., Bonn.