INTERNATIONAL WORKSHOP | APRIL 10–12 2024

# PERSPECTIVES ON HYBRID HUMAN–AI SYSTEMS

## Bringing Together Interdisciplinary Approaches

**Invited Speakers**

Christoph Bareither | Anne Dippel
Steven Dorrestijn | Orit Halpern
Catholijn Jonker | Gertraud Koch
Pietro Michelucci | Rainer Mühlhoff
Albrecht Schmidt | Roanne van Voorst

VENUE: CARL FRIEDRICH VON SIEMENS FOUNDATION | SÜDLICHES SCHLOSSRONDELL 23 | MUNICH
CONTACT: HHAIS-WORKSHOP@EKWEE.UNI-MUENCHEN.DE

JOHANNES MOSER | LIBUŠE HANNAH VEPŘEK | LEONIE THAL
INSTITUTE FOR EUROPEAN ETHNOLOGY AND CULTURAL ANALYSIS

LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN
Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation
Fritz Thyssen Stiftung für Wissenschaftsförderung
Carl Friedrich von Siemens Stiftung
MÜNCHNER VEREINIGUNG FÜR VOLKSKUNDE E.V.

# WELCOME

## To the Workshop

The research fields of *Human Computation* and *Hybrid Intelligence* investigate how the capabilities of humans and computers can be combined in novel ways, thereby overcoming the limitations of today's strictly computational *Artificial Intelligence systems*, and yielding the ability to address problems that neither can solve on their own.

Such hybrid human–AI systems are of interest to several scientific disciplines, including computer science, philosophy, sociology, design, and cultural anthropology, as they not only enable new computational and engineering capabilities while posing interesting new problems on their own but also influence societies at a structural level and our everyday life at the micro level. They elicit questions on the human–AI relations in these hybrid systems that are ethical, judicial, social, cultural, logical, algorithmic, practical, and material in nature.

This workshop brings together scholars, experts, and interested parties with an interest in hybrid human–AI systems to learn how different disciplines understand and approach these systems. The aim is to discuss how each discipline can contribute to an interdisciplinary understanding of the collaboration of humans and machines in such hybrid systems. With this workshop, we want to spark new conversations between different disciplines, addressing, among others, the following questions:

- How does Human Computation and Hybrid Intelligence relate to pursuits of and discourses on *Artificial General Intelligence / strong AI*?
- How can we grasp human-technology and societal relations unfolding within these systems?
- How is trust established and (re)negotiated in such systems?
- What forms of ethics arise in and with hybrid human–AI systems?
- What are the unique perspectives of the different disciplines, and how can these inform one another?
- How can we arrive at understandings that consider the disciplines' different approaches and modes of reasoning?
- How can we work together on these topics?

The workshop is part of the research project "Playing *in the Loop*: New Human-Software Relations in Human Computation Systems and their Impacts on the Spheres of Everyday Life" funded by the German Research Foundation. The Fritz Thyssen Foundation, the Carl Friedrich von Siemens Foundation, and the Münchner Vereinigung für Volkskunde e.V. provide additional funding for the workshop.

We wish you an exciting and stimulating event.

Johannes Moser, Libuše Hannah Vepřek, and Leonie Thal

# PROGRAM

## Wednesday, 10th of April

| | |
|---|---|
| **10:00–10:30** | **Registration & Arrival – Coffee** |
| **10:30–11:15** | **Johannes Moser & Libuše Hannah Vepřek**<br>*Workshop Opening* |
| **11:15–12:15** | **Pietro Michelucci**<br>*The Conundrum of Purpose in Collective Intelligence Systems* |
| **12:15–13:45** | **Lunch – Restaurant "Schlosswirtschaft Schwaige"** |
| **13:45–14:45** | **Gertraud Koch**<br>*Hybrid Intelligence - Reflecting AI Design Approaches through the Lens of Anthropology of Technology* |
| **14:45–15:00** | **Short Break** |
| **15:00–16:00** | **Orit Halpern**<br>*The Neural Doctrine: The Future of the Human in Machine Networks* |
| **16:00–16:30** | **Coffee Break** |
| **16:30–17:30** | **Designing & Working Together – Lightning Talk Session I**<br><br>**Cecilia Colloseus & Andrea Sell** – *AI-Cockpit – An Interdisciplinary Project on "Human in Command"*<br><br>**Tim Cech** – *Towards De-Anthropomorphization of AI in the Field of XAI*<br><br>**Matthias Hirth** – *Illustrating the Benefits of Human-AI Collaboration Using the Example of Identifying Cars in the Context of Autonomous Driving* |
| **17:30-17:45** | **Short Break** |
| **17:45-18:45** | **Catholijn Jonker**<br>*A Self-Reflective Perspective on Hybrid Human-AI Systems* |
| **18:45** | **Reception – Carl Friedrich von Siemens Foundation** |

# PROGRAM

## Thursday, 11th of April

| | |
|---|---|
| **9:00–9:15** | **Coffee** |
| **9:15–10:15** | **Learning, Understanding, & Questioning in Hybrid Human–AI Systems – Lightning Talk Session II**<br><br>**~~Isperih Marinov Karaivanov~~** ~~– *Application and Utility of Actor-Network Theory in Understanding the Implementation of Artificial Intelligence in Informal Learning and Formal Education Networks*~~<br><br>**Thorsten Neischwander** – *Capturing Historical Imagination Empirically through Deep-Learning Machines?*<br><br>**Richard David-Rus & Elena Popa** – *Understanding as Perspective Taking in the Context of Artificial Intelligence* |
| **10:15–10:45** | **Coffee Break** |
| **10:45–11:45** | **Rainer Mühlhoff**<br>*"Human-Aided AI" as a Philosophical Approach in the Ethics and Critique of AI* |
| **11:45–12:00** | **Short Break** |
| **12:00–13:00** | **Christoph Bareither**<br>*Hybrid Epistemic Practices: Generative Artificial Intelligence and the Transformation of Academic Assemblages in the Qualitative Social Sciences and Humanities* |
| **13:00–14:30** | **Lunch – Carl Friedrich von Siemens Foundation** |
| **14:30–15:30** | **Anne Dippel**<br>*Temporality and Thingnes(s)': What Do We Learn from How We Are (Working), When Humans and Machine Learning Are Getting Entangled* |
| **15:30–16:00** | **Coffee Break** |
| **16:00–17:00** | **Steven Dorrestijn**<br>*AI Life Coaching: The Hybrid Self Between Discipline and Self-Care* |
| **17:00–18:00** | **Break** |
| **18:00** | **Conference Dinner – Restaurant "Schneider Bräuhaus im Tal"** |

# PROGRAM

## Friday, 12th of April

| | |
|---|---|
| **9:00–9:15** | **Coffee** |
| **9:15–10:15** | **Roanne van Voorst**<br>*Researching Human-Nonhuman Collaborations, Now and in the Future* |
| **10:15–10:30** | **Short Break** |
| **10:30–11:30** | **Collaborative Futures? Exploring Hybrid Human–AI's Impacts in Healthcare, Information, & Work – Lightning Talk Session III**<br><br>**Maren Heibges, Marvin Kopka, Markus Feufel, Christine Schmid** –<br>*Large Language Models in Healthcare: Perspectives from Situated Ergonomics*<br><br>**Oshri Bar-Gil** – *Redefining Human-Centered AI: The Human Impact of AI-Based Recommendation Engines*<br><br>**Sabine Pfeiffer, Amelie Tihlarik, Dennis Eckhardt & Marco Blank** – *Giving Employees a Voice in Designing AI. Participatory Design Research on Hybrid Human-AI Systems Using "Denkzeug"* |
| **11:30–12:00** | **Coffee Break** |
| **12:00–13:00** | **Albrecht Schmidt**<br>*Will Artificial Intelligence Help Humans Gain Superpowers?* |
| **13:00** | **Closing** |

# ABSTRACTS

## Wednesday, 10th of April

**11:15–12:15**

**PIETRO MICHELUCCI**
Human Computation Institute / Cornell University

***The Conundrum of Purpose in Collective Intelligence Systems***

A concoction of unprecedented influence is being mixed before our very eyes, and along with artificial intelligence and augmented reality, humans are among the ingredients. When mixing a new potion, there are two questions to be answered: 1) what is its intended purpose?; and 2) what will be its actual effects?

It is often easier to answer the former than the latter, but today there is no consensus on the purpose of emerging thinking technologies. Indeed, we have found ourselves in a laboratory full of newly discovered reagents, and like mad scientists, we are mixing them with wild abandon in hopes of conjuring an elixir that will save the world. But even the notion of a “saved world” is both subjective and elusive, and not everyone is inclined to look beyond the enticement of immediate and personal gains. If we wish, as individuals and collectives, to realize our destinies, we must answer both questions. Yet we are ill-equipped to answer either.

So how can we proceed? Perhaps the stoic philosophers would have us sit back and allow events to unfold as we keep our minds in harmony with the universe. Would such an approach make sense when our collective thinking may represent a coalescence of intentionality that could itself be harnessed to define and engineer the future of everything? What if sipping the potion could help us answer these questions. What if, in a bootstrapping fashion, a first generation hybrid intelligence system could itself be used to help define the purpose of future such systems and anticipate their impacts. Indeed, it is precisely within the purview of such systems to tackle both problems: achieving consensus and modeling the future. Thus, leveraging the complementary capabilities of humans and machines may be our best hope for querying the space of possible futures toward designing a destiny of our collective volition.

# ABSTRACTS

## Wednesday, 10th of April

**13:45–14:45**

**GERTRAUD KOCH**
University of Hamburg

***Hybrid Intelligence - Reflecting AI Design Approaches through the Lens of Anthropology of Technology***

In the current debates about Artificial Intelligence (AI) design approaches gained specific attention. Hybrid intelligence, as a design approach to AI, emphasizes the different contributions of human and artificial intelligence to the system, thus favoring a 'weak definition' of AI that neither competes with humans nor pretends to be the more powerful intelligence. They propose augmenting the human intellect with a collaborative, adaptive, responsible and accountable artificial intelligence (Zeynep et al. 2020). Looking from inside a hybrid-intelligence AI project, conducted 2021-2024 together with the Language Technology Group Hamburg for setting up the D-Wise Toolsuite for multimodal digital discourse analyses, the image crumbles that the hybrid intelligence design approach is something separate or distinct from ‘strong AI approaches’. Moreover, design paradigms such as 'hybrid intelligence approaches', 'AI for social good', etc. follow normative and imaginative causes without having a direct translation into the complexities and contingencies of social reality.

Through the lens of the anthropology of technology, value-oriented AI design approaches need to be problematised. They are contested and confused by the motivations and interests of multiple stakeholders involved in the development of AI from basic research to implementation. These complexities of both AI design and implementation in diverse social contexts are poorly reflected in current approaches to AI design. The large language model (LLM), LifeLanguage, serves as an example to outline this complexity and reduce it to five design layers relevant to the emergence of AI LLMs. The layers open up enquiries into the complexity of the emergence of AI technology as a social fact and call for consideration of different stages and stakeholders of AI development, thus serving as entry points for broadening the scope of design possibilities and needs of AI.

In the tradition of design anthropology (Smith et al. 2016), the layers of design also offer a way of thinking about AI beyond intelligence, but as emergent in various technological activities, with many imaginative grounds for envisioning alternative AI design options and technological futures.

# ABSTRACTS

## Wednesday, 10th of April

**15:00–16:00**

**ORIT HALPERN**
Technical University Dresden

***The Neural Doctrine: The Future of the Human in Machine Networks***

This talk interrogates the history of models of decision making and agency in machine learning, neo-liberal economic thought, and finance in order to interrogate how reactionary politics, population, and technology are being reformulated in our present. In our present, with the public emergence of ChatGPT 4 and similar technologies, this question of how we are imagining “human” agency and intelligence is critical. While the relationship between the Right, post-truth, suggestion algorithms, and social media has long been documented, rarely has there been extensive investigation of how ideas of choice and freedom become recast in a manner amenable to machine automation and to particular brands of post-1970s alt-Right discourses. An analysis of this history demonstrates a new logic within algorithmic and artificial intelligent rationalities that intersects with, but is also not merely a repetition of, earlier histories of eugenics, sexism, and racism. This situation provokes serious challenges to political action, but also to our theorization of histories of race and sex capitalism.

# ABSTRACTS

Wednesday, 10th of April

**16:30–17:30**

**LIGHTNING TALK SESSION I:**
**DESIGNING & WORKING TOGETHER**

**Cecilia Colloseus & Andrea Sell**
Aalen University

***AI-Cockpit – An Interdisciplinary Project on "Human in Command"***

The research project "AI-Cockpit" develops a technology for an AI system that enables humans to maintain both an overall view and control of the AI's automatic decisions. The aim is to generate workflows that are simpler, safer and faster. In three different scopes of application (human resources, smart city, care), initial prototypes are being designed to provide overarching insights for the individual areas, and an open source software will be developed.
We take a people-centered approach using participatory research methods. When designing technological innovations, socio-technical research is of great importance. Our initial task as empirical cultural scientists on the interdisciplinary team will be to conduct a field-research at a care home for older adults. This will allow us to understand everyday work in the care home and figure out the employees' and residents' acceptance of technical innovations. Together with software engineers, a chatbot for care documentation is to be developed on the basis of the data gathered both in our field research and by quantitative data provided by the care home. One of the great challenges will be to bring together the cognitive interests and perspectives, the methodological and theoretical approaches of all researchers who are involved in the project in a goal-oriented way.
With our contribution we will present our initial experiences gathered in the project that allow us to draw some first conclusions about scientific enculturation processes and the opportunities that come with them.

# ABSTRACTS

## Wednesday, 10th of April

**16:30–17:30**

**LIGHTNING TALK SESSION I:**
**DESIGNING & WORKING TOGETHER**

**Tim Cech**
University of Potsdam

***Towards De-Anthropomorphization of AI in the Field of XAI***

Trust and understanding are essential requirements for human-agnostic Explainable AI (XAI), especially for black box models. However, trust and understanding are not generally quantifiable measures and thus are inconsistent with the quantitative approaches of current XAI methods (such as LIME (Ribeiro et al. 2016), Melody (Chan et al. 2020), and Grad-CAM (Selvaraju et al. 2017). Often, AI models are either trivial or incomprehensible for qualitative assessment. We propose that it is required to first investigate an AI model as a cultural product to fully understand the technical object. Thus, we expect that the quiddity of an AI model changes in different contexts, i.e., if it is investigated by individual users with different requirements, expectations, and evolving knowledge about the context itself.
First, we determine how state-of-the-art XAI techniques can be used to establish credible justification for the quality of outputs. For example, how we can create a credible 2D layout with dimensionality reduction.
Second, we devise methods that enable users to code their current understanding of the context in a quality metric, such as statistical hypothesis testing for a 2D layout.
Third, we want to enable users to iteratively optimize and change their quality metric until satisfying their needs using XAI methods.
We envision users to qualitatively refine outputs of XAI methods in correspondence with their evolving knowledge about their context. By doing this, we facilitate that their XAI method provides context-specific quantitative explanations. Thus, we can achieve de-anthropomorphization for XAI methods by avoiding the expectation that XAI methods can create a qualitative understanding of a context by themselves.

# ABSTRACTS

## Wednesday, 10th of April

**16:30–17:30**

**LIGHTNING TALK SESSION I:**
**DESIGNING & WORKING TOGETHER**

**Matthias Hirth**
Technical University of Ilmenau

***Illustrating the Benefits of Human-AI Collaboration Using the Example of Identifying Cars in the Context of Autonomous Driving***

Large, carefully curated training sets are necessary for artificial intelligence (AI) models to operate successfully and consistently. However, training data is scarce since its production typically requires manual expert annotation, which limits scalability.
This obstacle can be overcome with the use of crowdsourced microtasking, which provides access to a worldwide workforce and may allow for the cost- and time-efficient, highly scalable annotation of visual data. However, compared to completely automated methods, human work is still slower and more expensive. In this talk, we present a workflow based on human-AI cooperation that enables extensive, high-quality annotations of visual data for autonomous driving systems, notably identification automobiles. Three scenarios, each reflecting a distinct traffic and weather scenario, are used to evaluate the viability of this collaboration.
We found that crowdworkers improved the AI's work by identifying more than 40% of the missing cars. We do note that an automatic pre-annotation can assist the crowdworkers, though. In this scenario, a human-AI collaboration increases the annotation quality and annotation efficiency.

# ABSTRACTS

## Wednesday, 10th of April

**17:45-18:45**

**CATHOLIJN JONKER**
Delft University of Technology

***A Self-Reflective Perspetive on Hybrid Human-AI Systems***

In this talk we show our advancements in combining knowledge technology with black box AI systems, to increase the monitoring of black box AI systems. Based on these advancements we propose the use of Knowledge Technology in communication with humans to formulate and maintain ethical properties for monitoring both the input and output stream of data of AI systems. The Hybrid Intelligent system consisting of Human and Knowledge-based AI keep the underlying Black Box AI system under meaningful human control.

# ABSTRACTS

Thursday, 11th of April

**9:15–10:15**

**LIGHTNING TALK SESSION II:**
**LEARNING, UNDERSTANDING, & QUESTIONING IN HYBRID HUMAN–AI SYSTEMS**

**Isperih Marinov Karaivanov**
Firelink Consult

***Application and Utility of Actor-Network Theory in Understanding the Implementation of Artificial Intelligence in Informal Learning and Formal Education Networks***

As an analytical practice actor-network theory (ANT) is tailor-made to uncover and map the multitude of underlying relationships and effects exercised among the human and non-human actors in emerging complex systems such as human/AI hybrids. ANT has previously been criticized due to its perspective on viewing and accommodating said actors as forming heterogeneous networks whereby they are allowed equal agency, independent of their human or nonhuman status. However, with the advent of practices of literally offloading decision-making or learning regulation to the AI component in a hybrid system, such criticisms seemingly fall by the wayside. Nevertheless, ANT's actual use in studies on these topics is low at present and its potential thus is largely left unexplored.
The purpose of this presentation is to acquaint those unfamiliar with the ANT methodology and how it can be modified in order to be adopted within their own research parameters. The discussion will focus on both the conceptual and practical advantages and disadvantages of the approach, and will utilize as an example case study a hypothetical learning actor-network as derived from some of the currently available literature on hybrid human-AI learning practices, exploring the changes that result from the introduction of an AI within said network. It is important to emphasize that in order to make best use of ANT – in itself predominantly a descriptive tool – one is almost expected to combine it with another theoretical approach focused on the interpretative side of the research.

# ABSTRACTS

## Thursday, 11th of April

**9:15–10:15**

**LIGHTNING TALK SESSION II:**
**LEARNING, UNDERSTANDING, & QUESTIONING IN HYBRID HUMAN–AI SYSTEMS**

**Thorsten Neischwander**
Philipps-University Marburg

***Capturing Historical Imagination Empirically through Deep-Learning Machines?***

The presentation introduces the planned doctoral project "Capturing historical imagination empirically through deep-learning machines" in the field of didactics of history. The aim is to diagnose the historical imagination of learners with the help of image-generating AI systems and to shed light on further questions of historical learning, such as questions about the ideas of history that exist in a society.
Alongside historical culture and historical consciousness, historical imagination is one of the core concepts of the didactics of history for describing historical learning. In contrast to the other categories, however, imagination is more difficult to capture empirically and has been less researched to date.
As part of the doctorate, a method will therefore be tested to enable teachers to gain insights into the imagination of their students. This can not only have an influence on the planning of lessons but can also reveal ideas about history in a society. The lecture will discuss the collaboration between humans and machines in classroom research and hopes for impulses from other disciplines for the further development of the project.

# ABSTRACTS

## Thursday, 11th of April

**9:15–10:15**

**LIGHTNING TALK SESSION II :**
**LEARNING, UNDERSTANDING, & QUESTIONING IN HYBRID HUMAN–AI SYSTEMS**

**Elena Popa,** Jagiellonian University Krakow **& Richard David-Rus,** Romanian Academy

***Understanding as Perspective Taking in the Context of Artificial Intelligence***

Current literature on explainable AI emphasizes a diverse set of explanatory needs given the multiple stakeholders involved in the AI ecosystem. This results in a range of explanation-based understanding processes and products that are not fully inter-commensurable. Among others, this has ethical consequences, such as blindspots regarding patterns of epistemic injustice or biases in algorithms reinforcing discrimination. We argue that humanistic understanding in the form of perspective taking can help alleviate such vulnerabilities. We discuss perspective taking as situating oneself in the web of relationships of the target subject or seeing the world in terms of how it affords actions and satisfies the interests of the target persons. We show that perspective taking can be mapped onto explanatory understanding in terms of knowledge of aims and goals as defined by von Wright. We further argue that objectual understanding, defined by Wilkenfeld as being able to manipulate its representation in the right sort of way, can address shortcomings of using von Wright's model, notably correct mistaken representations.

# ABSTRACTS

## Thursday, 11th of April

**10:45–11:45**

**RAINER MÜHLHOFF**
University of Osnabrück

***"Human-Aided AI" as a Philosophical Approach in the Ethics and Critique of AI***

In this talk, I will present the concept of 'Human-Aided AI' and the critical paradigm in the ethics of AI that it enables. Human-Aided AI is a perspective on data-driven and machine learning AI systems that recognizes the continuous dependence of such systems on human input as an essential part of the technology. AI systems are then sociotechnical systems; artificial intelligence capacities are not solely computed in computing farms but are products of hybrid brain- and silicon-based computations. In a genealogical approach, I will point out that the breakthrough of Deep Learning was only possible because of media-cultural developments that enabled the massive capture of human cognitive skills and continues to rely on this media infrastructure today. I will then highlight that discrimination of users, social sorting, and predictive analytics are inherent capabilities of Human-Aided AI systems, which has important consequences in the ethics and critique of AI.

# ABSTRACTS

## Thursday, 11th of April

**12:00–13:00**

**CHRISTOPH BAREITHER**
University of Tübingen

***Hybrid Epistemic Practices: Generative Artificial Intelligence and the Transformation of Academic Assemblages in the Qualitative Social Sciences and Humanities***

Generative artificial intelligence (AI), such as OpenAI's Generative Pre-Trained Transformer (GPT), is rapidly transforming many areas of everyday life. One of the areas severely affected is academic research and teaching. Universities around the world are struggling to make sense of the newfound power of generative AI. Not only can students now easily produce academic texts, numerical calculations, or code with a few simple “prompts”; senior researchers are also starting to apply generative AI in their epistemic practices. Ethnographic research guided by digital anthropology and STS approaches can make a significant contribution to the study of these transformations. This presentation introduces a theoretical framework based on "assemblage thinking" that can help to study how "academic assemblages" are transformed through the emergence of "hybrid epistemic practices" - that is, epistemic practices that emerge through the collaboration between human actors and AI systems.

# ABSTRACTS

## Thursday, 11th of April

**14:30–15:30**

**ANNE DIPPEL**
University of Jena

***Temporality and Thingnes(s)': What Do We Learn from How We Are (Working), When Humans and Machine Learning Are Getting Entangled***

Multiple agents weave the fabric of temporality in a data-driven world, at the same time orchestrating futures and visions of (un)certainty and sparking promises to contain what is to come. The possibilities of modeling the world with algorithmic means are fundamentally changing perceptions of humans experiencing time and space - new publics, new socialities, new identities are emerging, and rules of coexistence must be negotiated with them. The times they are-a-manyworlding.
What interests me in this talk, is how phenomena come into being and how temporality operates as part of research processes within infrastructures, where humans and computation are entangled. Hardware, software and algorithms in a techno scientific as well as posthuman world of many worlds are producing new entities and agencies on planet Gaia.
In this talk, I will be discussing the concept of intra-viduality - developed out of empirical research in the world of physics and theoretical engagement with queer-feminist natural philosophy. Thereby, I hope to contribute anthropological insights to the investigation of hybrid human-AI systems. I will be thus presenting research from a field where machine learning is deployed since the 1980s: high-energy-physics at CERN (Switzerland).
The discussion of these borderlands of hybrid human computation will be based on data gathered during a long-term qualitative research in algorithmic worlds focussing on entanglements of hardware, software and intra-actions of group-related work dynamics. I will aim to look into the rhythms of “doing things” by various agents (human and non-human) and the ethical challenges emerging out of the operational realist mode of understanding the world on the one hand and the more-than-semiotic realities of humans-cum-thingness’ agencies on the other. By investigating material cuts and infrastructural ruptures which disturb and destabilise seemingly seamless work environments, I hope to shed light on the question whose decision will matter and what role power and hierarchies have in future worlds of “human computation". As outlook I intend to pose the question what role Anthropology as discipline might play in working out what is to come.

# ABSTRACTS

## Thursday, 11th of April

**16:00–17:00**

**STEVEN DORRESTIJN**
Saxion University of Applied Sciences

***AI Life Coaching: The Hybrid Self Between Discipline and Self-Care***

Developments in AI make technology interfere with human existence in new ways. This presentation brings up three cases of ‘AI for life coaching’. These will be analyzed in a framework inspired by Michel Foucault with hybrid self, care of the self and disciplinary power.
1) Since it became widely available end of 2022 ChatGPT has made big impression. A chatbot that you can conversate with in human language, which not only helps you searching information, but helps you processing the information and which writes a conclusion for you. This is not just a neutral means that you employ deliberately, but it takes over your activities and your deliberations, it interferes with your way of living and of being.
2) This effect is even more apparent in the case of the ‘anger wearable’ for people with problems with aggression. The appliance monitors and interprets heartbeat, sweat, skin tension, breathing, in order to help someone recognizing if he/she is getting angry and help avoiding an outburst of aggression.
3) Ultimately this could lead to the fictious case of ‘Gina’, the AI assistant which/who is the main character in a novel by Ewoud Kieft. The Gina continuously accompanies a person, as a practical help but also as a life coach, becoming a kind of technical conscience and therapist. The fascinating plot shows how such a device can be both a technology for care of the self and a vehicle for disciplinary power.

# ABSTRACTS

## Friday, 12th of April

**9:15–10:15**

**ROANNE VAN VOORST**
University of Amsterdam

***Researching Human-Nonhuman Collaborations, Now and in the Future***

With health data being considered countries' 'future oil', public and scholarly concerns about 'algorithmic ethics' rise. Research has long shown that algorithmic datasets (re)produce social biases, discriminate and limit personal autonomy. At the same time, it is also clear that alghorithms are able to improve public health.
This presentation introduces a research project that takes place in six different countries around the world, investigating human and nonhuman collaboration in hospitals. It asks how medical experts are influenced by algorithms in their decisionmaking, and what are the underlying cultural and social factors that impact this collaboration.
While the research springs from anthropological investigations, it includes an experimental 'future in the now' methodolology which is developed by researchers, future-foresighters, medical experts and programmers.

# ABSTRACTS

Friday, 12th of April

**10:30–11:30**

**LIGHTNING TALK SESSION III:**
**COLLABORATIVE FUTURES? EXPLORING HYBRID HUMAN–AI'S IMPACTS IN HEALTHCARE, INFORMATION, & WORK**

**Maren Heibges, Marvin Kopka, Markus Feufel, Christine Schmid**
Technical University of Berlin

***Large Language Models in Healthcare: Perspectives from Situated Ergonomics***

Our research group at Technische Universität Berlin specialises in Human Factors and Ergonomics, an area known as "Arbeitswissenschaft" in German, within the healthcare sector. Our work primarily explores decision-making practices in the medical field, with a particular emphasis on the digitalisation of healthcare systems. Our research adopts an interdisciplinary approach that merges perspectives and methodologies from cognitive and social sciences (emphasising practice-theory and situatedness), as well as engineering. This amalgamation distinguishes us from classic ergonomic research rooted predominantly in psychology or engineering (hence "situated" ergonomics). Our team, comprising psychologists, anthropologists, and engineers, closely collaborates with physicians and patients, enabling us to conduct both quantitative experimental research and ethnography, that is, various forms of mixed methods.
For the forthcoming workshop on "Hybrid Human-AI Systems", we intend to explore conceptual, practical, and social implications of Large Language Models (LLMs) in healthcare. LLMs, such as ChatGPT, are pre-trained artificial neural networks. Previous systems were predominantly designed for either physicians (with differential diagnoses) or patients (with self-diagnosis suggestions). However, LLMs may be used to address both use cases, offering a versatile tool for clinicians and patients alike. We would like to contribute from three perspectives, offering insights from both applied and theoretical dimensions that align with our group's research focus.

- Conceptual: We aim to delineate what LLMs in healthcare do on a conceptual level and what kind of data they use to generate results.
- Practical: We will outline how patients might use LLMs for medical consultation and what challenges this might produce, presenting some of our experimental data.
- Social(-theoretical): Building upon our understanding of social theory, we intend to critically examine the training data of these models. We aim to discuss potential biases and also touch upon the more philosophical question, what it means for social theory (particularly practice-theory and material-semiotics) that LLMs are purely language-based.

# ABSTRACTS

Friday, 12th of April

**10:30–11:30**

**LIGHTNING TALK SESSION III:**
**COLLABORATIVE FUTURES? EXPLORING HYBRID HUMAN–AI'S IMPACTS IN HEALTHCARE, INFORMATION, & WORK**

**Oshri Bar-Gil**
IDF Behavioral Science Research Institute

***Redefining Human-Centered AI: The Human Impact of AI-Based Recommendation Engines***

In recent years, the flood of information has become overwhelming, and we experience constant information overload. Information overload is not new nor unique to our age. It is already well documented in the classical works of Simon (1971) with technological solutions offered even earlier (Bush, 1945). Some affordances in artificial intelligence (AI), specifically machine learning (ML) algorithms and big-data analytics promise to augment the user's information selection, processing, and decision making to make a less stressful, "frictionless" life (Andrejevic, 2013). Typically, these solutions combine ML algorithms, such as recommendation engines (for information search, news, music, shopping, etc.), and apply them to the digital data double of the user (a digital data model designed to reflect a human being, also called data-doppelgänger). Zuboff (2018) claimed that most platforms use this model, named surveillance capitalism, to acquire large profits. While doing so, the technology companies construct the political, sociological, and economical spheres to fit the surveillance capitalism model. Although discussion of those implications is of immense importance, it is ignoring the question at the title of this article – Can AI-based information processing algorithms, aimed to augment and even replace our internal cognitive processing, be considered human-centered AI?
To answer this question, the chapter will survey the effects of using recommendation engines based on data- doppelgängers on the users' sense of self. The chapter will use prior research to focus on four key areas of influence on users' self-perception: autonomy, agency, rationality, and memory (Bar-Gil, 2020). It will suggest that although these technologies might afford new capabilities to augment human users' engagement in the information sphere, they have a considerable effect on the users – diminishing their agency, making them less autonomous, more confirmative to algorithmic rationality, resulting in changes to their memory patterns.

# ABSTRACTS

Friday, 12th of April

**10:30–11:30**

**LIGHTNING TALK SESSION III:**
**COLLABORATIVE FUTURES? EXPLORING HYBRID HUMAN–AI'S IMPACTS IN HEALTHCARE, INFORMATION, & WORK**

**Sabine Pfeiffer, Amelie Tihlarik, Dennis Eckhardt & Marco Blank**
Friedrich-Alexander-University Erlangen-Nürnberg

***Giving Employees a Voice in Designing AI. Participatory Design Research on Hybrid Human-AI Systems Using "Denkzeug"***

The integration of AI technologies into daily life, from voice assistants to purchase recommendations, has be come pervasive. ChatGPT, in particular, has gained widespread adoption in Germany, with one in four individuals using it (TÜV/Forsa 2023). This phenomenon offers a unique opportunity to explore the evolving dynamics of human-AI interaction. The expectations of AI systems replacing human tasks, be they routine or creative, are high. However, research in the workplace domain highlights the undervalued competencies of employees and the overestimated capabilities of AI in replicating human performance (Pfeiffer 2020).

Still too often, the importance of incorporating employees' contextual and experiential knowledge from the outset in designing robust AI systems and hybrid human-AI collaborations is underscored. Early participation of employees not only empowers them to contribute to AI system design but also leverages their rich organizational insights, often overlooked in AI design processes (Herrmann/Pfeiffer 2023).

Our innovative approach, "Denkzeug," employs a modified Q-Method (Johnson/Waishwell 2014) to quantify subjective perspectives on hybrid collaboration between AI and employees. This tool, available in both online and offline versions, stimulates reflection and discourse on human-centric AI design. It covers 16 statements on AI in the workplace along 8 dimensions related to hybrid collaboration, shedding light on employees' perceptions (Pfeiffer 2023).

We present "Denkzeug" and its application in on- and offline surveys as well as in participative research within our "Labouratory" research (Sauer et al. 2021), and central findings. Our work contributes to advancing the dialogue on meaningful and human-friendly AI design, emphasizing the invaluable perspective of employees in shaping the future of work.

# ABSTRACTS

## Friday, 12th of April

**12:00–13:00**

**ALBRECHT SCHMIDT**
LMU Munich

***Will Artificial Intelligence Help Humans Gain Superpowers?***

The quest to enhance and expand human capabilities has been a catalyst for innovation throughout history. Human evolution and development are tied to tools and technologies that make us stronger, faster, healthier, and smarter. Technological and methodological inventions have made people more effective in their interactions with each other and their environment. Technology has expanded our range of communication, and we have developed machines that can literally move mountains. As humans learned to read and write, and with the Internet, human knowledge became widely available and enhanced human cognitive abilities.

How will recent advances in artificial intelligence (AI) push this to new limits? Will human creativity and ingenuity be enhanced by generative AI and multimodal foundation models? Will this take human augmentation to a whole new level? Or is this the beginning of a massive cognitive decline?

In our research, we experimentally explore digital technologies for non-invasive human augmentation and enhancement, with a focus on seamless interaction with intelligent tools. Users should no longer be aware that they are receiving technical assistance. The enhanced and augmented capabilities should feel natural, and control should be implicit and not add to their cognitive load. It should simply feel like getting a superpower. With ample opportunity to collect data in the wild and the means to create large language and foundational models, we may have the ingredients to develop hybrid human-AI systems. Are we just one step away from providing compelling superpowers to enhance the physical and cognitive capabilities of individuals and society?

# VENUES

## How To Get There

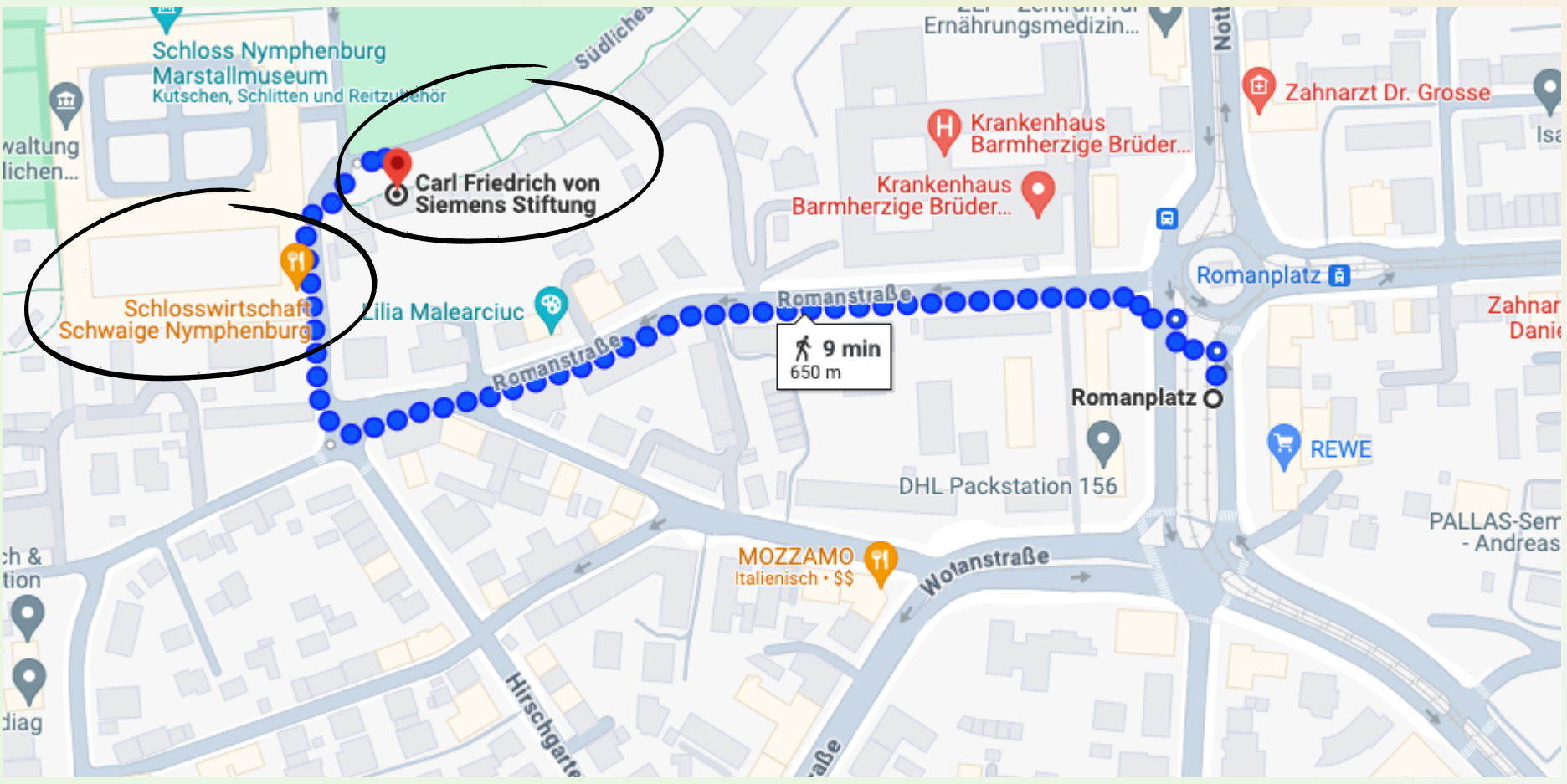
**WORKSHOP VENUE**

*Carl Friedrich von Siemens Foundation*
*Südliches Schlossrondell 23 | 80638 München*

You will find the foundation at Südliches Schlossrondell 23. Walk towards the side wing of Nymphenburg Palace, it is in the last yellow “Kavaliershaus” on the left.
From Munich central station, take the U1 in the direction “Olympia-Einkaufszentrum” and disembark at “Rotkreuzplatz” (3 stops). Exit the subway station and follow the “Tram” signs. Take tram line 12 to “Romanplatz”. From there it is about 650 metres by foot via Romanstraße to Südliches Schlossrondell 23.
From Laim S-Bahn station, take bus line 51 or 151 to Romanplatz.
By car, please turn onto Südliche Auffahrtsallee in the direction of Schloss Nymphenburg. This leads directly to the palace and merges seamlessly into the Südliche Schlossrondell (one-way street). Please note: The Foundation does not have its own parking spaces.

**LUNCH ON WEDNESDAY, 10TH OF APRIL**

***Restaurant “Schlosswirtschaft Schwaige”***
*Schloß Nymphenburg 30 | 80638 München*



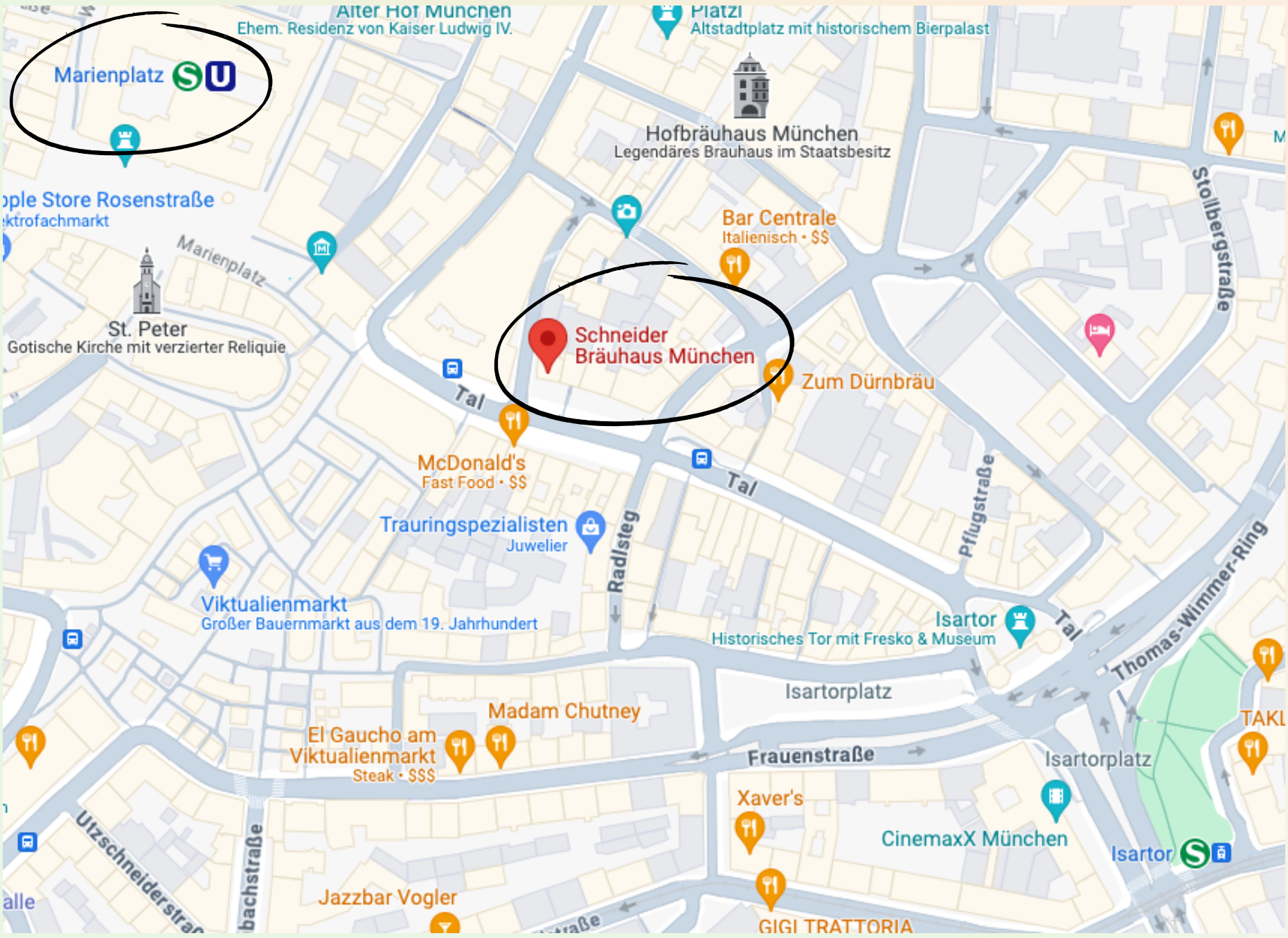
# VENUES

## How To Get There

**CONFERENCE DINNER ON THURSDAY, 11TH OF APRIL**

***Restaurant "Schneider Bräuhaus"***
*Tal 7 | 80331 München*

The restaurant is located in the city centre, a three minute walk from Munich's central "Marienplatz".
The best way to reach it is by public transport: Take the S-Bahn or U3 and U6 to "Marienplatz", follow the signs to the "Tal" exit. Bus 132 also stops at Marienplatz, the bus stop is right in front of the restaurant.



# TRAVEL INFORMATION

## How To Get Around

The best way to get around Munich is to use public transportation.

To look up connections and buy mobile tickets, it is best to use the **"MVGO"** app from the local public transport provider "MVG".
You can also use it to rent sharing bikes.
You can download it here:

**iOs**: https://apps.apple.com/de/app/mvgo/id991757585

**Android:** https://apps.apple.com/de/app/mvgo/id991757585

If you do not want to buy tickets on your phone, you can find ticket vending machines at all subway (U-Bahn) stations or inside buses and trams.
You only need tickets for zone M (one way ticket is €3,80, Single Day Ticket €9,20 and Group Day Ticket (2-5 adults) €17,80)

For further information on public transportation in Munich see: www.mvg.de