PUI Perceptual User Interfaces

University of Stuttgart Germany

OPEN THESES EVENT

Perceptual User Interfaces Group

Prof. Dr. Andreas Bulling

Summer Term 2024

Perceptual User Interfaces Group, University of Stuttgart

www.perceptualui.org

Distribution is strictly prohibited.

Our Group
















Table of Contents

Collaborative Intelligence

Artificial intelligent (AI) systems increasingly implement perceptual, learning, decision-making, and interactive skills

Collaborative Intelligence

Artificial intelligent (AI) systems increasingly implement perceptual, learning, decision-making, and interactive skills

Perception

Autonomous cars

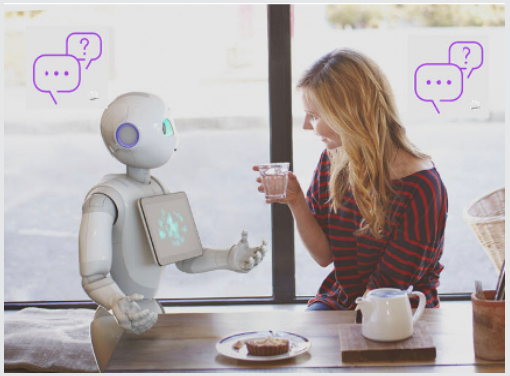
Collaborative Intelligence

Artificial intelligent (AI) systems increasingly implement perceptual, learning, decision-making, and interactive skills

Perception	Learning
	
Autonomous cars	Intelligent assistants

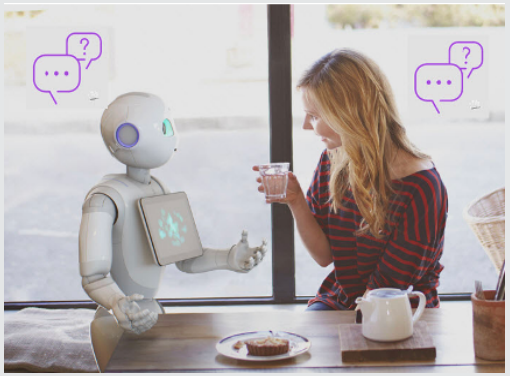
Collaborative Intelligence

Artificial intelligent (AI) systems increasingly implement perceptual, learning, decision-making, and interactive skills

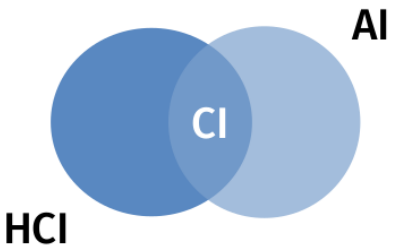
Perception	Learning	Interaction
		
Autonomous cars	Intelligent assistants	Social robots

Collaborative Intelligence

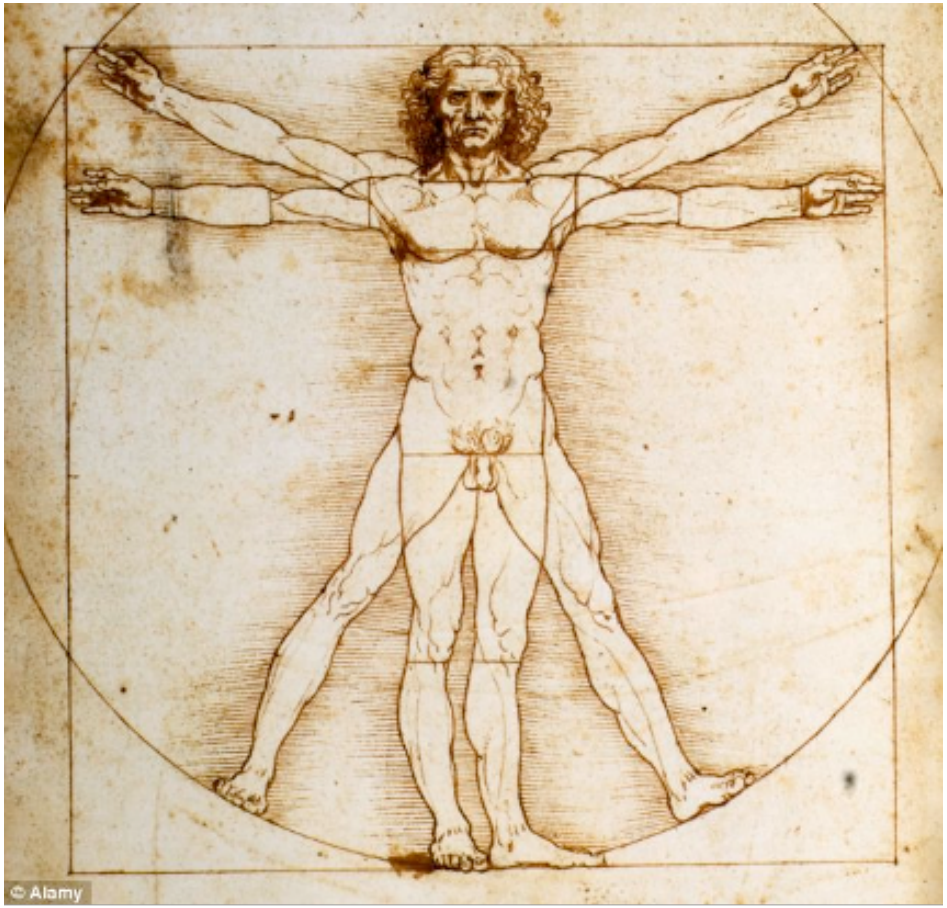
Artificial intelligent (AI) systems increasingly implement perceptual, learning, decision-making, and interactive skills

Perception	Learning	Interaction
		
Autonomous cars	Intelligent assistants	Social robots

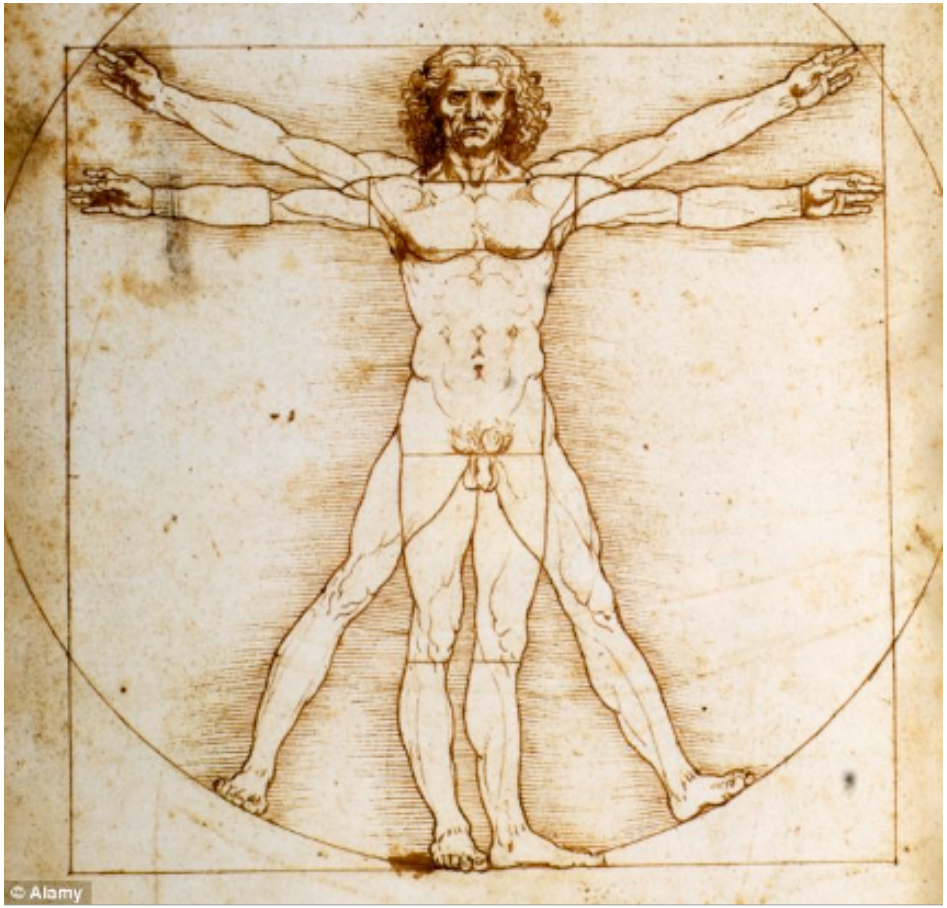
→ **Collaborative Intelligence**



Computational user modelling is key

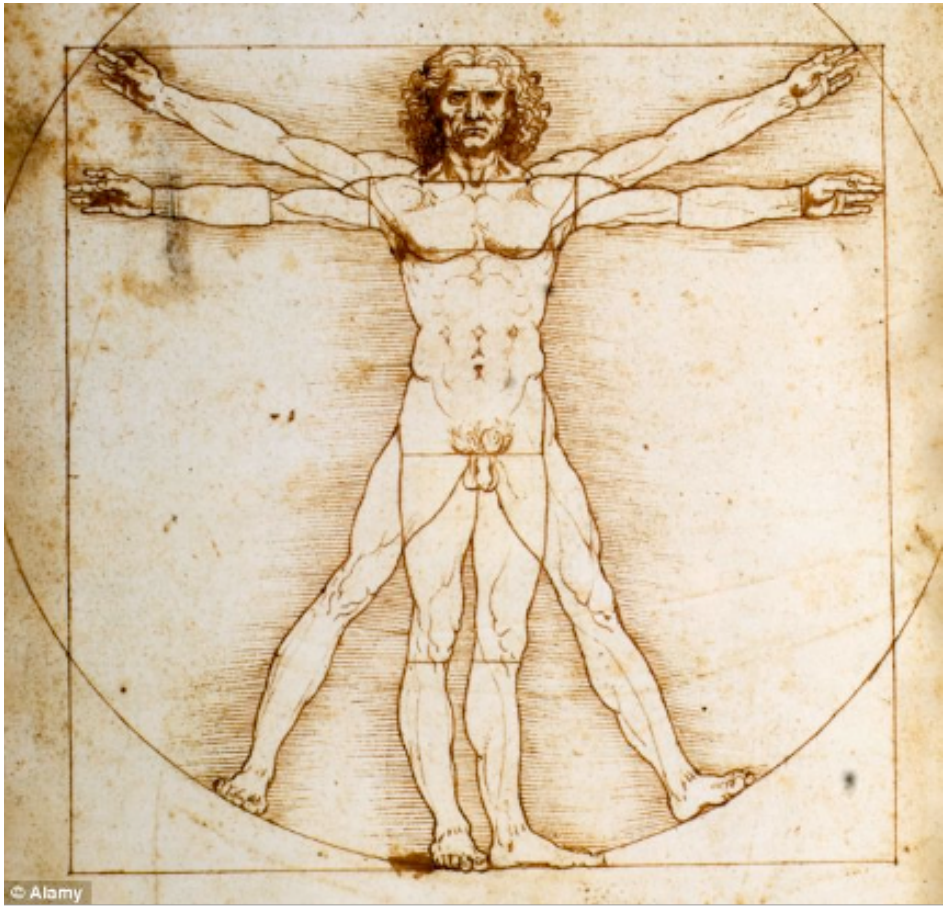


Computational user modelling is key



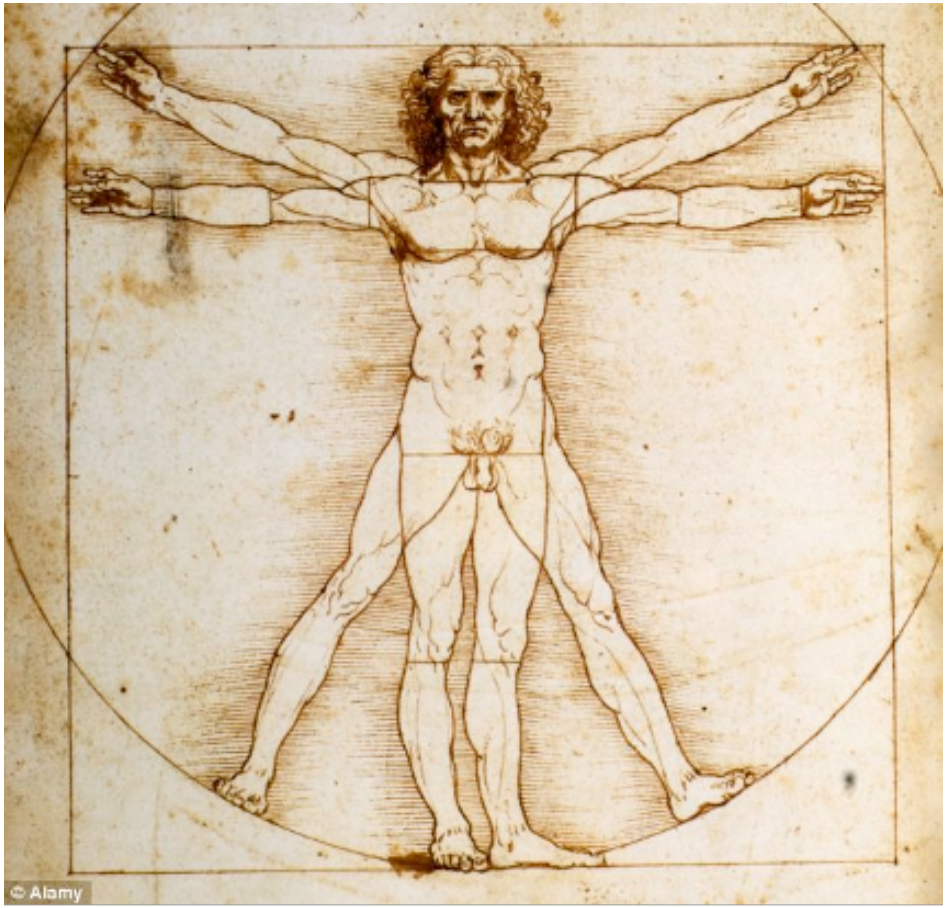
- Non-verbal signalling

Computational user modelling is key



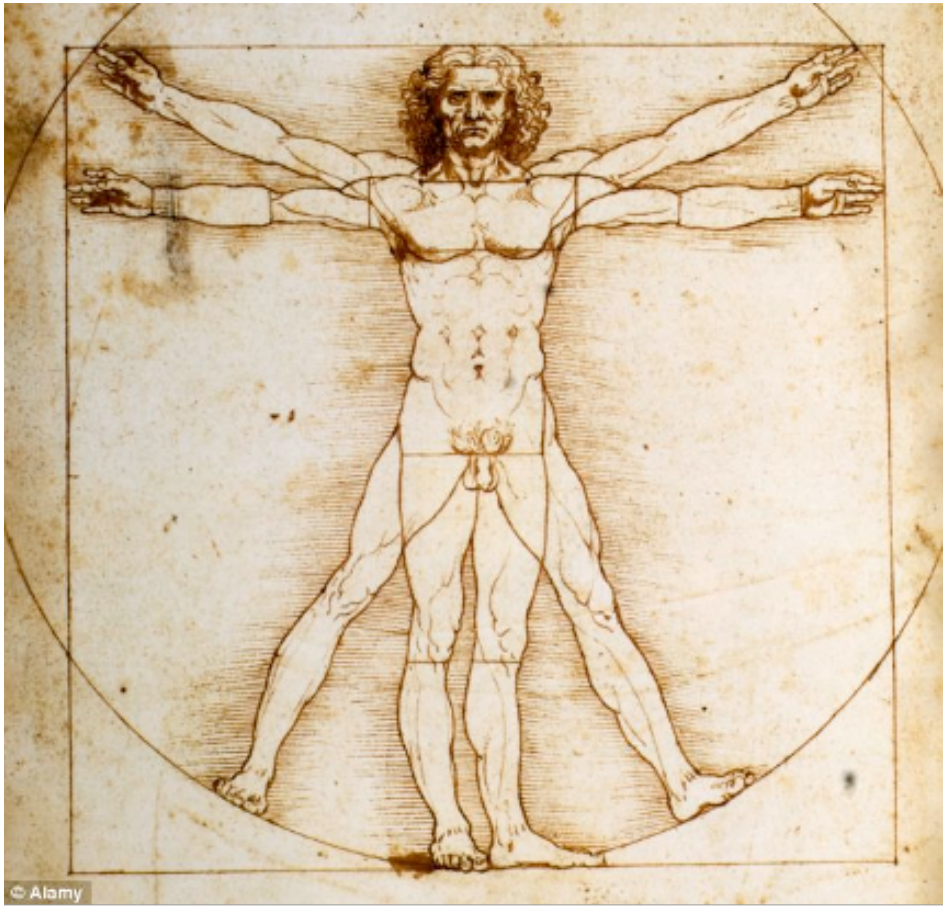
- Non-verbal signalling
- Cognition / Theory of Mind

Computational user modelling is key



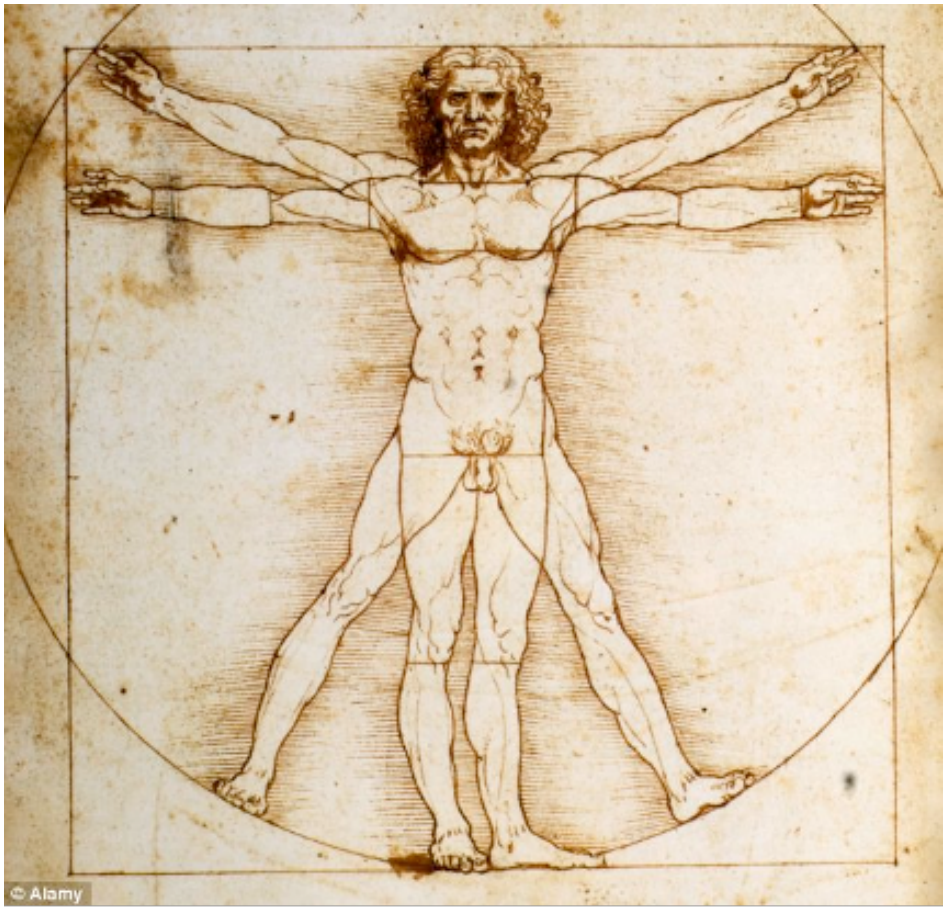
- Non-verbal signalling
- Cognition / Theory of Mind
- Physiological state

Computational user modelling is key



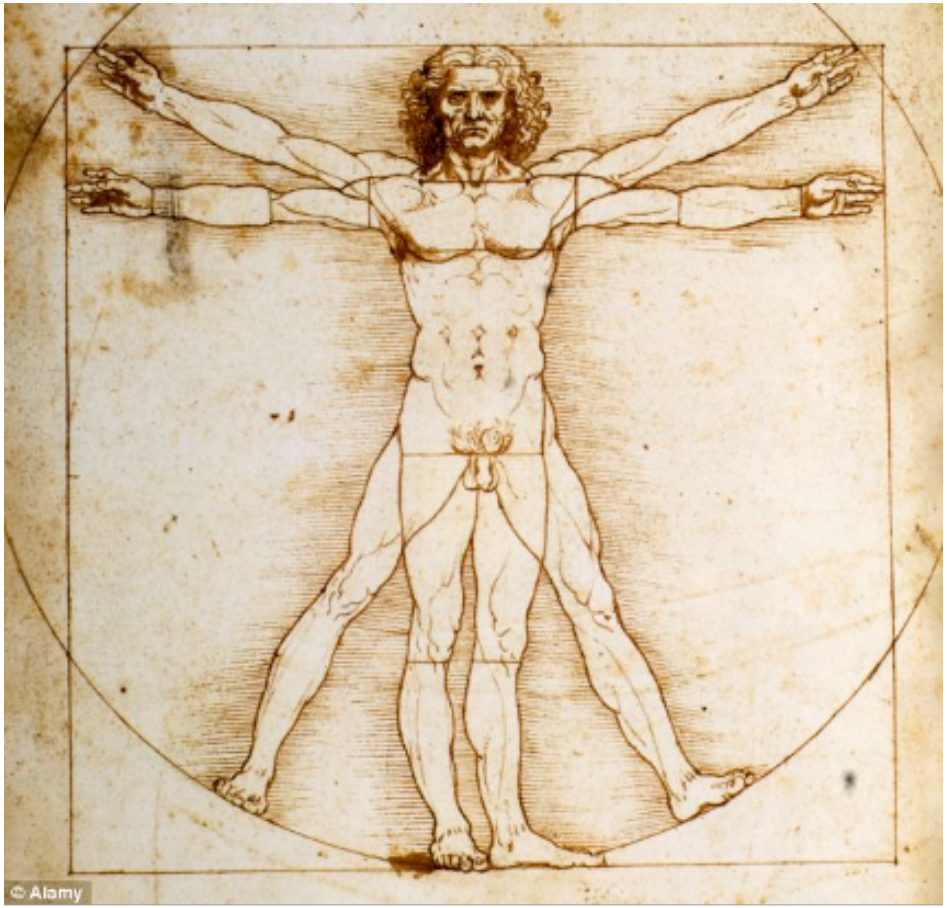
- Non-verbal signalling
- Cognition / Theory of Mind
- Physiological state
- Affective state

Computational user modelling is key



- Non-verbal signalling
- Cognition / Theory of Mind
- Physiological state
- Affective state
- Personality

Computational user modelling is key



- Non-verbal signalling
- Cognition / Theory of Mind
- Physiological state
- Affective state
- Personality
- Social behaviour

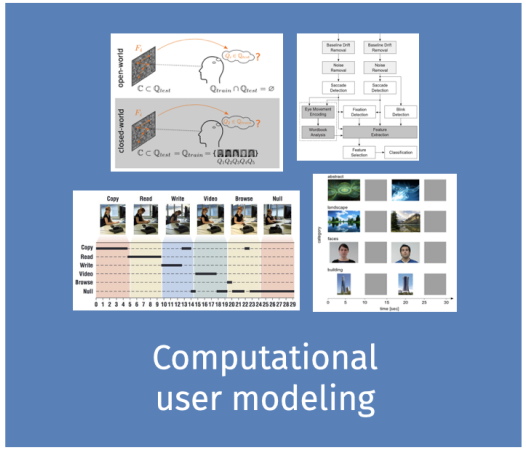
Research areas I

ACM UbiComp'08,'14; JAISE'09
Eurographics'15,'18; ACM CHI'13,'18,'19
ACM ETRA'12,'14,**'16,'18,'19**
ACM UIST'15,**'16,'17**
ICCV'15; IEEE CVPR'15; CVPRW'17
ECCV'16; PACM IMWUT'17
IEEE TPAMI'19; BMVC'20

Research areas I

Everyday user behaviour sensing

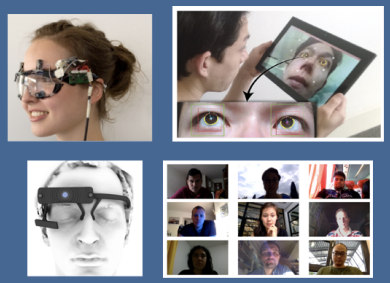
ACM UbiComp'08,'14; JAISE'09
Eurographics'15,'18; ACM CHI'13,'18,'19
ACM ETRA'12,'14,**'16,'18,'19**
ACM UIST'15,**'16,'17**
ICCV'15; IEEE CVPR'15; CVPRW'17
ECCV'16; PACM IMWUT'17
IEEE TPAMI'19; BMVC'20



Computational user modeling

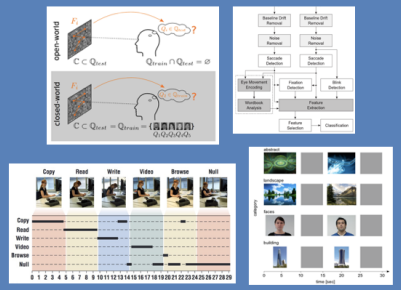
ACM UbiComp'09,'11,'15
IEEE TPAMI'11; IEEE PCM'11,'14,'20
ACM TAP'11; ACM CSUR'14
IEEE CVPR'15,'17; IEEE ICCVW'17
ACM CHI'13,**'16**,'20,'22
ACM MobileHCI'18
NeurIPS'20**;** CoNLL'20
Neurocomputing'20
ICCV'21; JOV'21; **ACM ETRA'19**,'22
IEEE TVCG'21,'22,'23

Research areas I



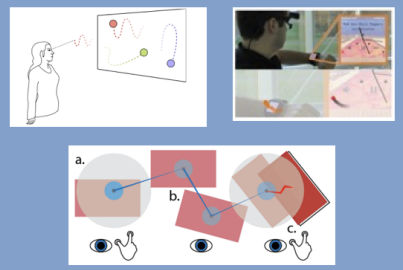
Everyday user behaviour sensing

ACM UbiComp'08,'14; JAISE'09
Eurographics'15,'18; ACM CHI'13,'18,'19
ACM ETRA'12,'14,**'16,'18,'19**
ACM UIST'15,**'16,'17**
ICCV'15; IEEE CVPR'15; CVPRW'17
ECCV'16; PACM IMWUT'17
IEEE TPAMI'19; BMVC'20



Computational user modeling

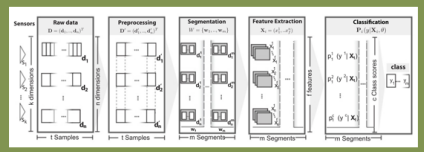
ACM UbiComp'09,'11,'15
IEEE TPAMI'11; IEEE PCM'11,'14,'20
ACM TAP'11; ACM CSUR'14
IEEE CVPR'15,'17; IEEE ICCVW'17
ACM CHI'13,**'16**,'20,'22
ACM MobileHCI'18
NeurIPS'20**;** CoNLL'20
Neurocomputing'20
ICCV'21; JOV'21; **ACM ETRA'19**,'22
IEEE TVCG'21,'22,'23



Multimodal interaction and dialog

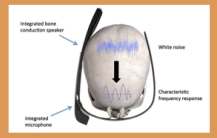
ACM CHI'12,'15,'16,'17,**'18**
ACM UbiComp'13,'14,'15,'16
ACM UIST'13,**'15**,'16,'17
Springer PUC'15,'17; ACM CSUR'15
IEEE Computer'16; **MUM'17**
ACM ETRA'20 (x2)
CoNLL'21
OzCHI'22; **COLING'22**

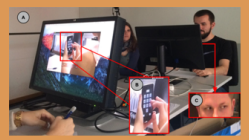
Research areas II



Computational human behaviour analysis

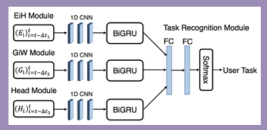
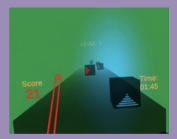
ACII'13,'15; ACM CHI'13
ACM CSUR'14
AH'13,'14; GCPR'14
ACM IUI'18; ACM ICMI'19,'21
ACM Multimedia'21,'22,'23



Usable security and privacy

ACM CHI'12,'16,'18, '23; ACM PerDis'17
ACM UbiComp'14; MobileHCI'15
ACM ICMI'17; **MUM**'16,**'17**
PACM IMWUT'18**;** ACM ETRA'20
PETs'22; NeurIPS GMML'22
IEEE PCM'23


Head-mounted augmented and virtual reality

IEEE VR'16; AVI'18
ACM CHI'19,'21
ACM ETRA'20
IEEE TVCG'21; ACM SSUI'21
ACM TOCHI'22

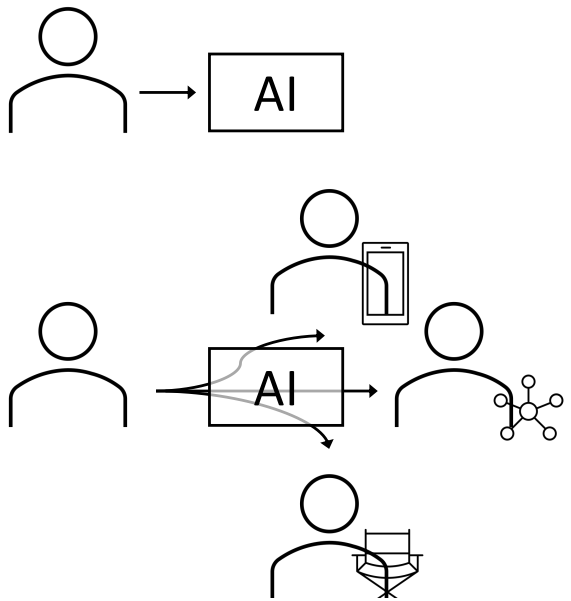
Table of Contents

Open Theses Topics

Susanne

Research Interests

- Explainable AI (XAI)
- Counteract anthropomorphism (treating AI systems like humans)
- Explanations about the human minds behind AI systems



Research Goal

Redirecting Theory of Mind for eXplainable AI (XAI) to understand the human minds behind AI systems.

Open Theses

Explaining Disagreement in VQA with Eye Tracking (BSc/MSc)

"What profession is the man on the right?"

"priest"

"drummer"

- Human annotation (annotation interface, filtering candidate cases, conducting annotation, dataset validation)
- Extension to other datasets: Byproducts Han et al. [2023], SalChartQA Wang et al. [2024] (co-supervised by Yao Wang)

Bring your own idea!

susanne.hindennach@vis.uni-stuttgart.de

Open Theses Topics

Zhiming

Research Interests

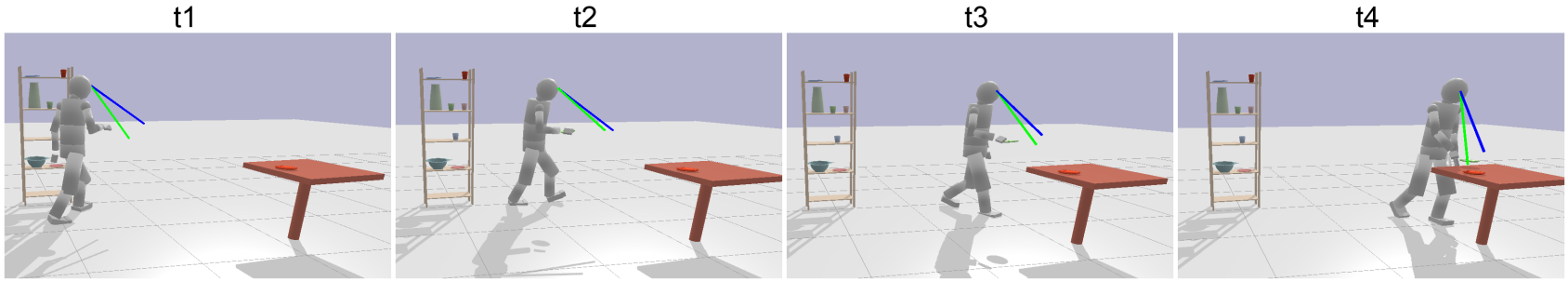
- Human-computer interaction
- Virtual reality
- Eye tracking
- Human-centred artificial intelligence

Research goal

Develop deep learning methods for modelling human behaviours in activities of daily living

Motivation

- Eye and body movements are correlated in daily activities
- Eye and body movements are influenced by the scene environment

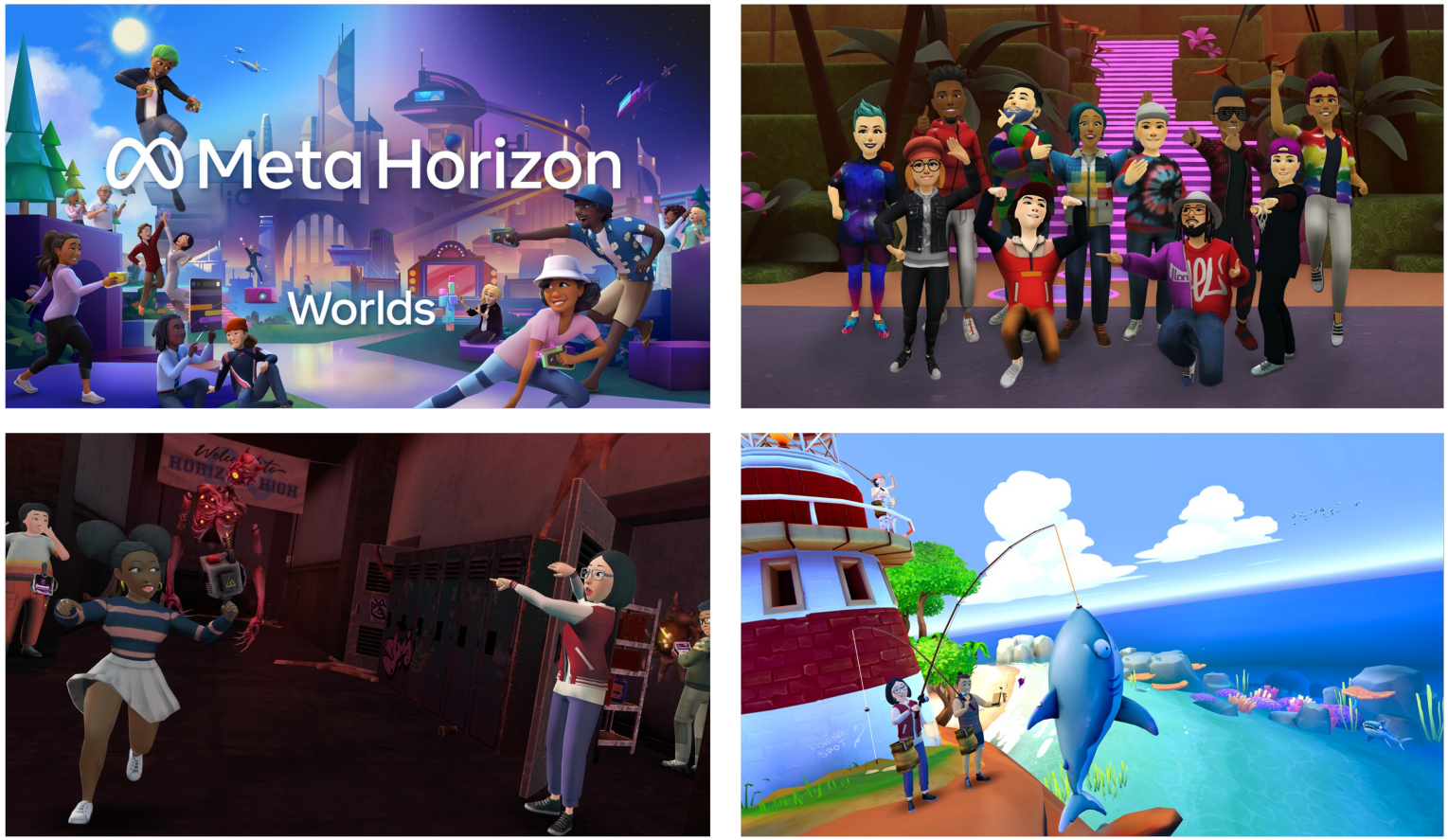


Source: Coordination of eye, body and scene environment

Motivation

Source: Virtual social interactions [Latoschik VRST '17]

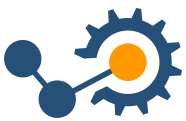
Motivation



Source: Large-scale virtual worlds [Meta]

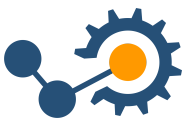
Optional Goals:

- Human intention estimation from eye gaze and body motions
- Human motion generation in human-human social interactions
- Joint prediction of human eye gaze and body motions
- Scene-aware human motion and eye gaze generation



Bring your own idea and contact me!

zhiming.hu@vis.uni-stuttgart.de



Open Theses Topics

Chuhan

Research Interests

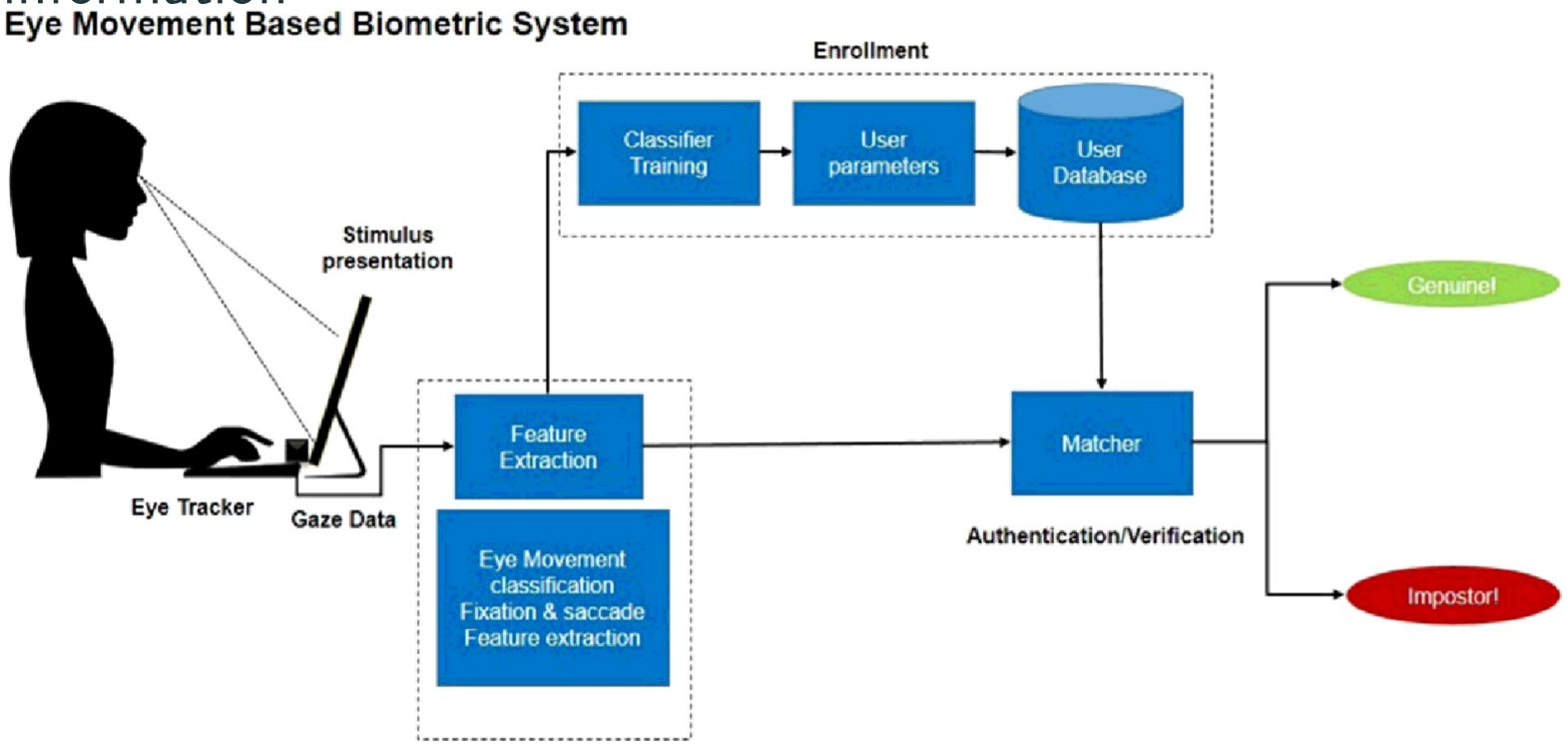
- Human-computer interaction
- Eye-tracking
- (3D) Computer Vision

Research goal

Representation Learning of Gaze Data

Eye tracking: Analyzing User-specific Information in High-frequency Eye Movements (Master)

- High-frequency human eye movements contain user-specific information



Eye tracking: Analyzing User-specific Information in High-frequency Eye Movements (Master)

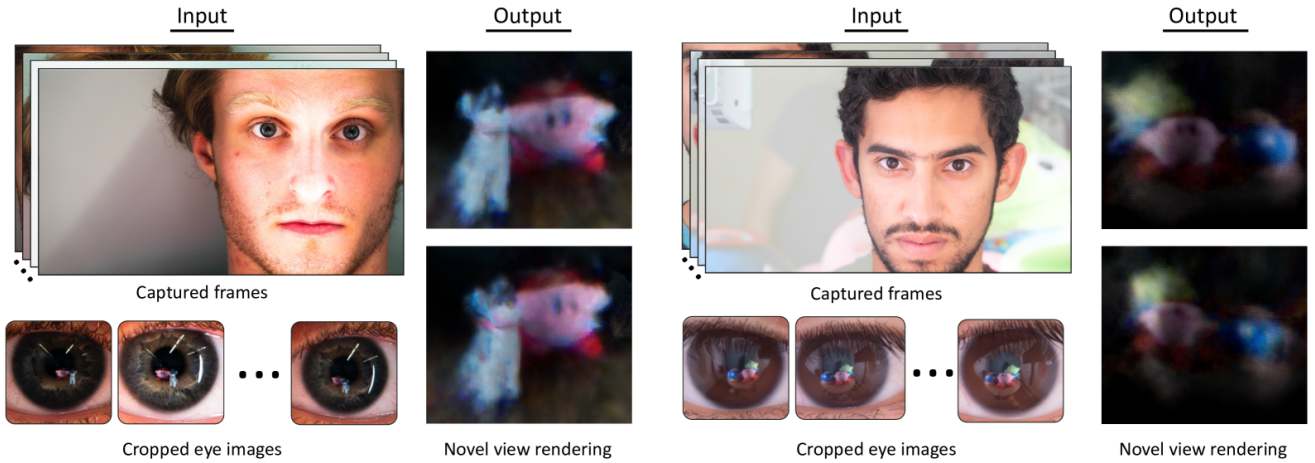
Which part of eye movements is related to human identity has not been explored.

Goal:

- Implement a deep learning model with frequency transformations for user authentication/identification
- Use existing XAI methods to analyze which frequency components in human gaze data is related to human identity

3D Computer Vision: Seeing the World through Your Eyes using 3D Gaussian Splatting (Master)

- Reconstruct the object you are looking at through the eye reflection.



Get in touch!

Important

All topics are available only after mid-June.

chuhan.jiao@vis.uni-stuttgart.de

Open Theses Topics

Constantin

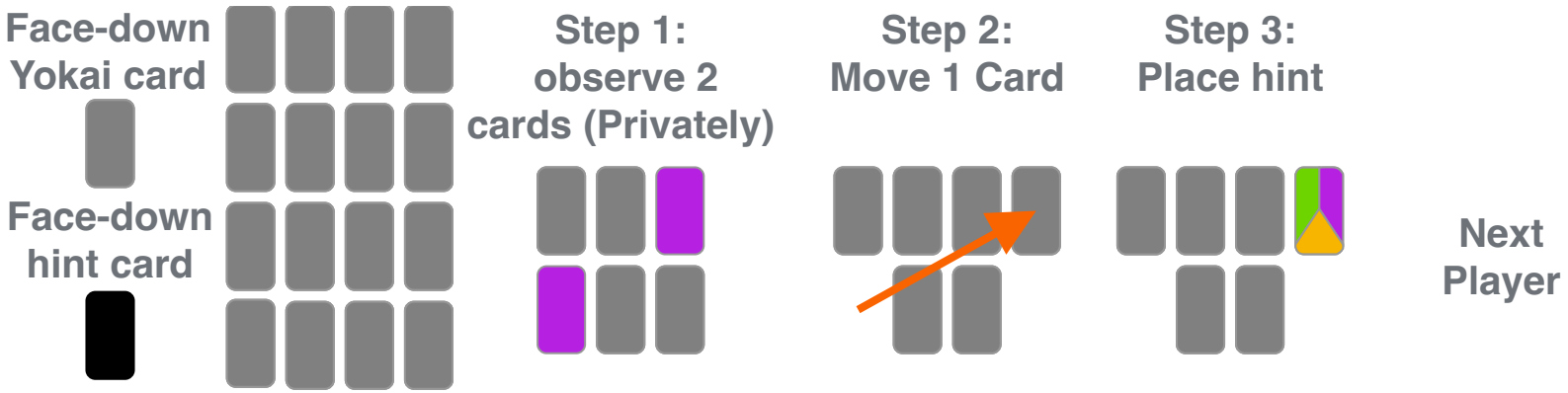
Research Interests – Learning To Collaborate With Humans

- Deep Multi-Agent Reinforcement Learning
- Human modeling
- Human-AI collaboration
- Computational Theory of Mind

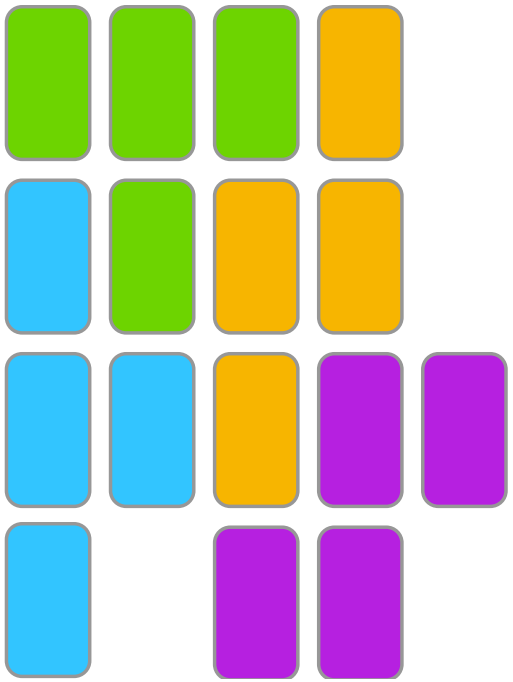
Research goal

Artificial agents capable of human-AI cooperation, enhanced with social reasoning capabilities.

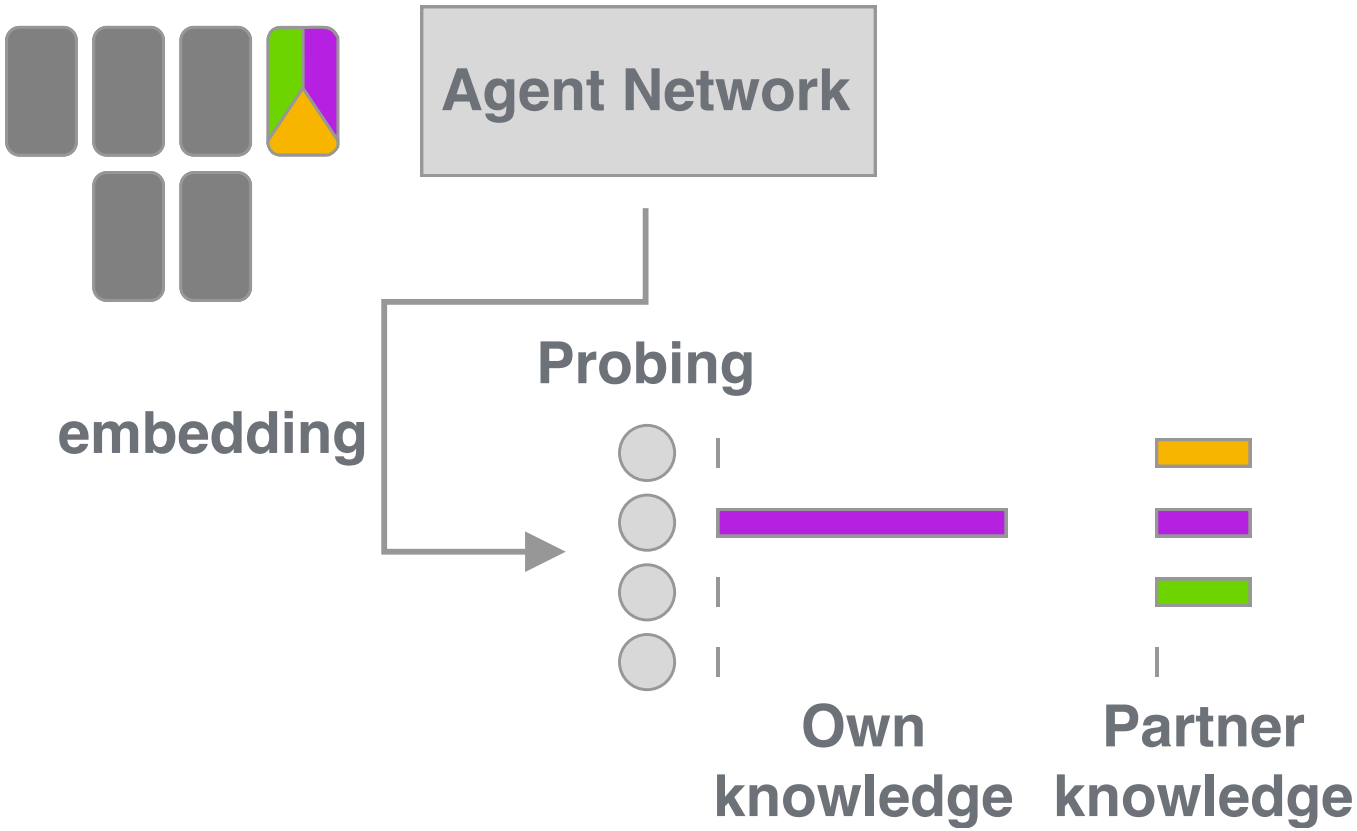
Probing for Theory of Mind in the Multi-Agent Cooperation Game Yokai



Winning Configuration



Task: Can LLMs or Reinforcement Learning Agents Represent Beliefs?



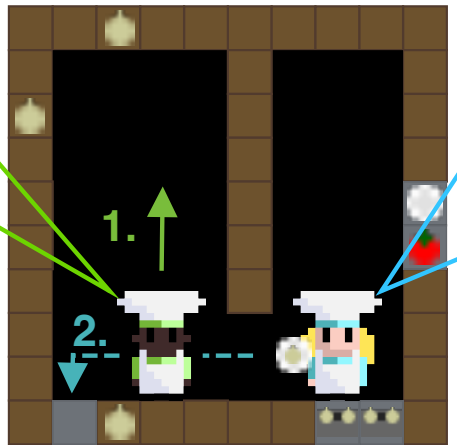
Probing for Theory of Mind in the Multi-Agent Cooperation Game Yokai

Tasks:

1. Built a text-interface for a Yokai Environment
2. Explore an LLMs ability to play Yokai
3. Hand design inputs for assessing LLM's Machine Theory of Mind Rabinowitz et al. [2018] capabilities
4. Explore the representations via Linear Probing (see Zhu et al. [2024]; Marks and Tegmark [2023])
5. Compare to results on text benchmarks Zhu et al. [2024]
6. Train RL Policies on the Yokai environment and compare findings

Partner Modeling in Deep MARL for Theory of Mind Tasks

I think he wants to deliver the soup ...



He needs to move out of the way, the soup is important!

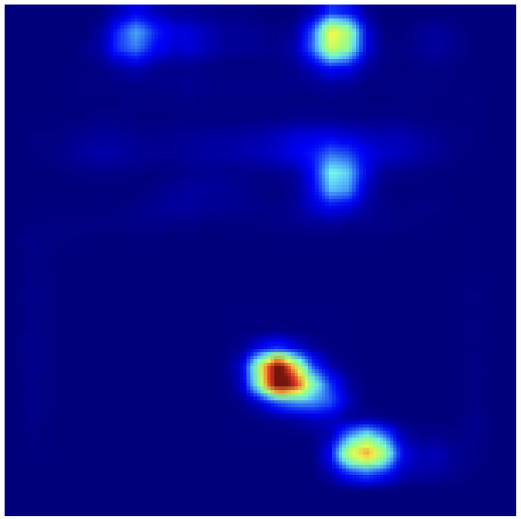
Partner Modeling in Deep MARL for Theory of Mind Tasks

Tasks:

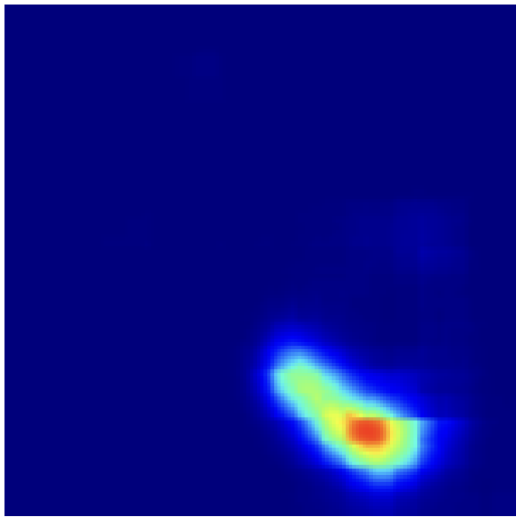
1. Survey the literature on opponent modeling He et al. [2016]; Chandrasekaran et al. [2017]; Raileanu et al. [2018]; Yu et al. [2022]
2. Decide on the feasibility of latent imagination based methods Hafner et al. [2020, 2021]; Samsami et al. [2024]
3. Test them on several Machine Theory of Mind Rabinowitz et al. [2018] benchmarks (at least 2)
4. Candidates: Hanabi Bard et al. [2020], Yokai, SymmToM Sclar et al. [2022]
5. Analysis (Linear Probing or come up with other)

Guiding Reinforcement Learning with a Cognitive Model of Human Visual Attention (Master) – with Anna Penzkofer

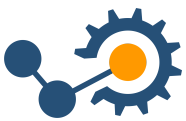
(a) Game State



(b) RL Attention



(c) Human Gaze



Guiding Reinforcement Learning with a Cognitive Model of Human Visual Attention (Master) – with Anna Penzkofer

1. Use a cognitive model (e.g. EMMA Salvucci [2001]) to simulate human attention in Atari games and compare to human attention (Atari-HEAD Zhang et al. [2020])
2. Guide imitation learning via CGL Saran et al. [2021]
3. Compare against other gaze guided imitation learning methods
4. Either show efficiency in a new environment without gaze (i.e. Overcooked, Melting Pot ...) or propose other interesting analysis

Contact me!

constantin.ruhdorfer@vis.uni-stuttgart.de

Open Theses Topics

Matteo

Research Interests

- Computational Theory of Mind
- Multi-agent collaboration
- Common-sense reasoning / intuitive psychology

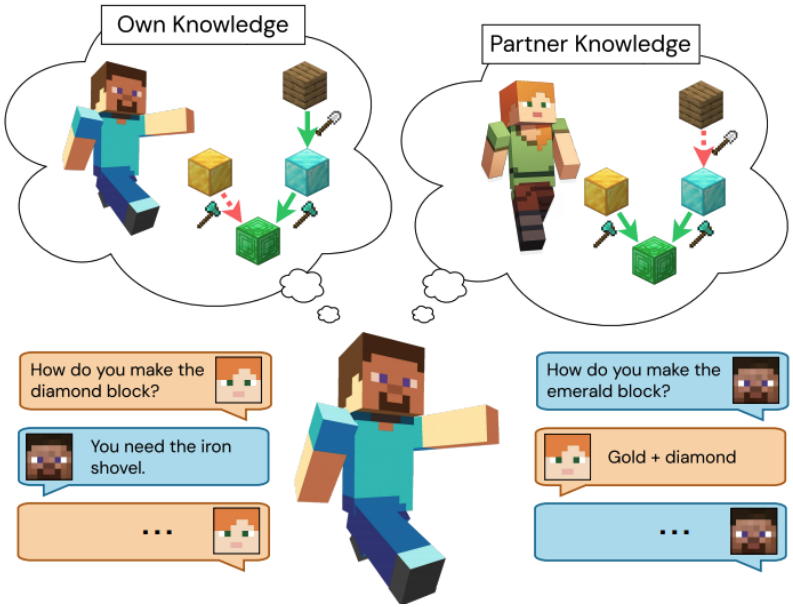
Research goal

Building AI systems that can infer mental states of others.

If you have an idea and you think I would think it is interesting, feel free to contact me!

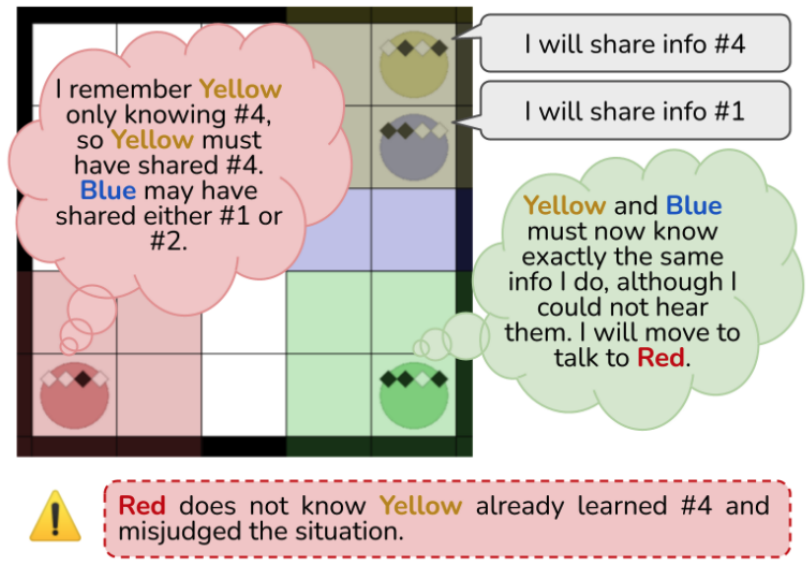
Theory of Mind in Collaborative Environments (Master)

- Theory of Mind plays a crucial role in collaboration
- Theory of Mind and Collaborative Plan Acquisition have been studied only from an observer point of view (MindCraft Bara et al. [2023]), without requiring to take actions
- Goal: Extend MindCraft to a multi-agent reinforcement learning environment



Theory of Mind Modelling in Symmetric Multi-Agent Environments (Master)

- SymmToM is an environment in which agents can speak, listen, see other agents, and move freely in a gridworld Sclar et al. [2022]
- Goal of the game: Collect all of the information available
- Goal of the thesis: Propose new methods for modelling Theory of Mind



Source: Sclar et al. [2022]

Open Theses Topics

Guanhua

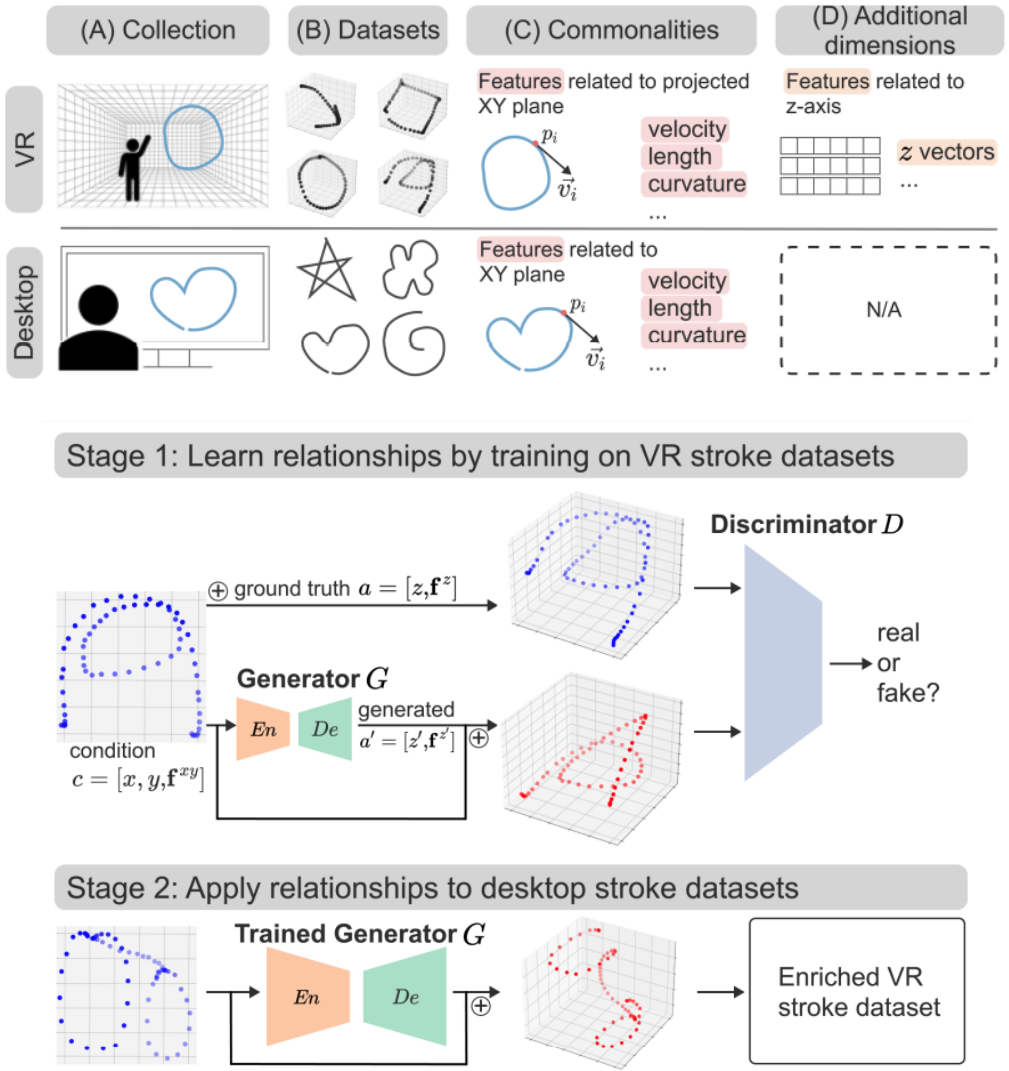
Research Interests

- Interactive behaviour and user interface modelling
- Representation learning

Research goal

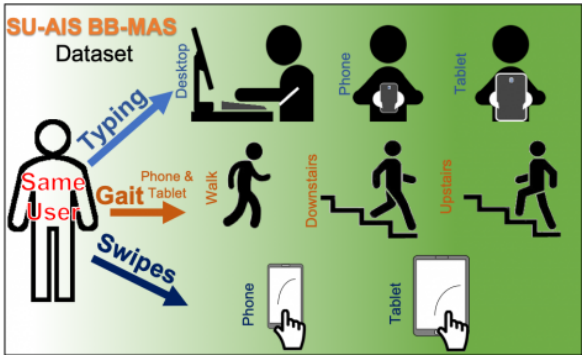
Semantic and reusable representations

Background – Cross-Device Behaviour Generation



Background – Existing Cross-Device Dataset

SU-AIS BB-MAS



Item		Description
Number of users		117
Activities		Typing, Gait and Swipes from ***the SAME USERS on all devices***
Devices (per participant)		Desktop, 2 Phones (hand and pocket), Tablet
Avg. and StD. (per user) of number of	keystrokes on Desktop	11760, 2132
	keystrokes on Phone	9415, 1463
	keystrokes on Tablet	8966, 1584
	keystrokes from a user on all devices	30153, 3880

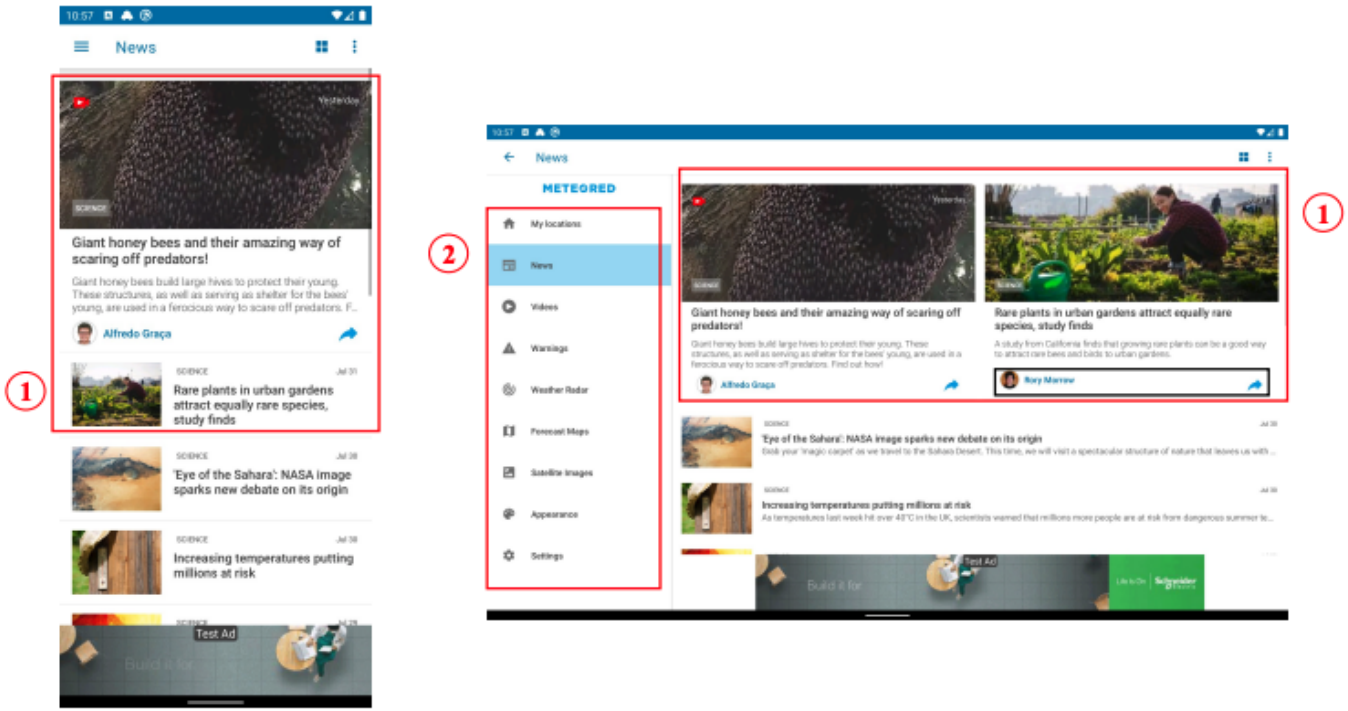
Belman et al., SU-AIS BB-MAS (Syracuse University and Assured Information Security - Behavioral Biometrics Multi-device and multi-Activity data from Same users) Dataset, IEEE Dataport, 2019

Master Thesis: Cross-Device Typing Behaviour Generation

1. Analyse the difference between desktop, phone and tablet devices
2. Simulate behaviour across these devices using generative models, e.g., diffusion
3. Evaluate the generated data
    - Keep the characteristics of this device and this user

Requirements: Statistical analysis, deep learning, Pytorch

Background – Cross-Device User Interfaces



UI only, missing interactive actions

Hu et al., Pairwise GUI Dataset Construction Between Android Phones and Tablets, NeurIPS'23 Track on Datasets and Benchmarks

Bachelor/Master/Research Project: Cross-Device UI&Action Dataset

1. Implement a recording system on mobile (phone and tablet) and/or desktop devices (GitHub codebases)
2. Recruit participants and collect user interfaces + interactive behaviour
3. Analyse the collected data: Statistics, behaviour patterns, usability, etc.

Requirements: Statistical analysis, Python

Great to have: Web development, Database

Bring Your Own Ideas

guanhua.zhang@vis.uni-stuttgart.de

Open Theses Topics

Yao

Research Interests

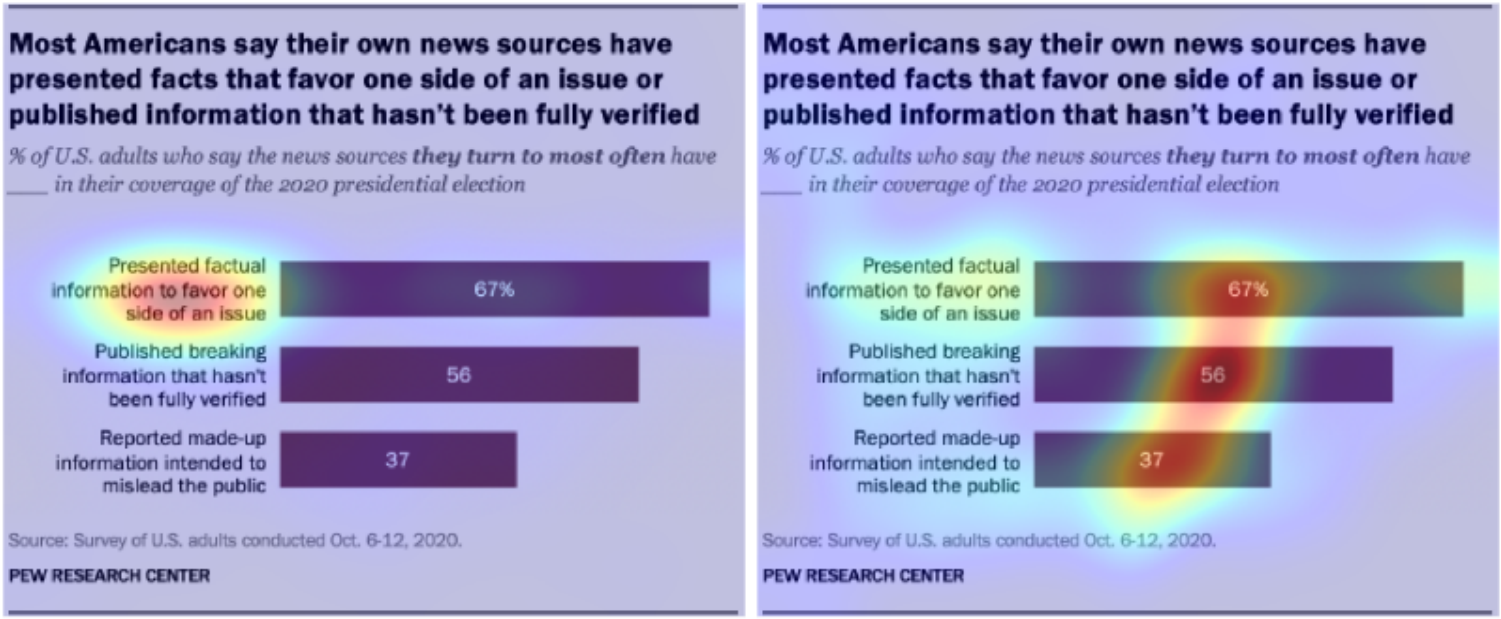
- Eye-tracking on information visualisations (2D, 3D)
- Computational models of human visual attention (saliency map, scanpath) under different tasks
- Metrics for evaluating visual scanpaths

Research Goal

Proposing good computational models of human visual attention to replace eye trackers in the production environment; optimising information visualisations from human visual attention

Analysis of the SalChartQA Dataset (Bachelor/Master)

Source: Left: Which reason has the highest amount of votes which is 67? Right: What is the average of all reasons?



- SalChartQA contains 3,000 visualisations and 6,000 question-driven visual saliency maps.
- Find more insights between gaze behaviour and some metrics, such as question correctness and answer uncertainty.

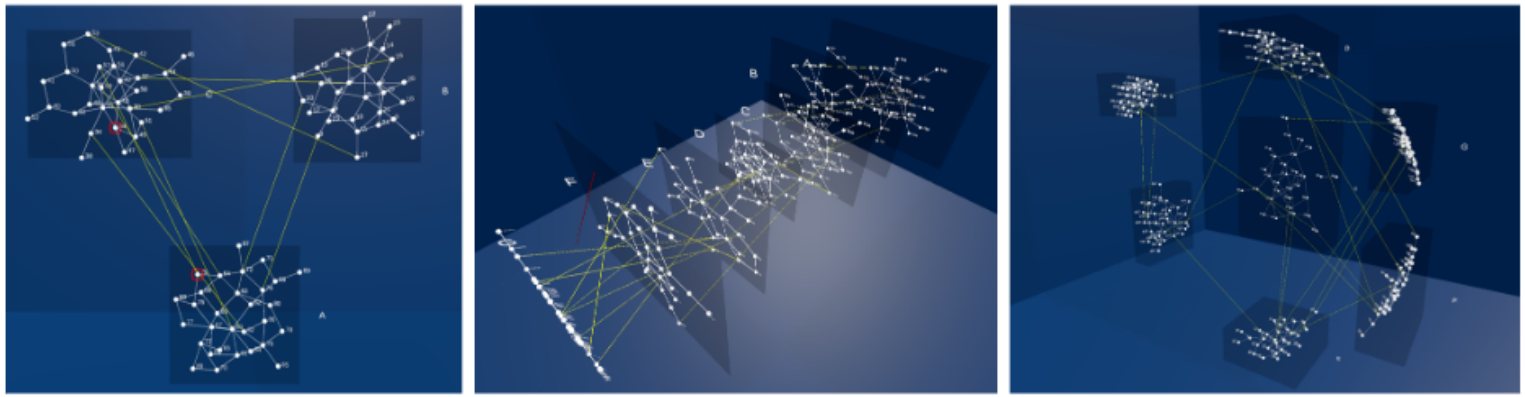
Example research questions:

1. Can gaze data represent the correctness / uncertainty of answers (how many participants answered this question correctly, how many unique answers are there)?

2. How do viewing patterns differ under different questions and different visualisation types, especially regarding areas of interest? Does the viewing patterns change over time (e.g. the first 3 seconds vs. the last 3 seconds)?

Understanding Visualisations in VR with Eye Tracking (Bachelor/Master)

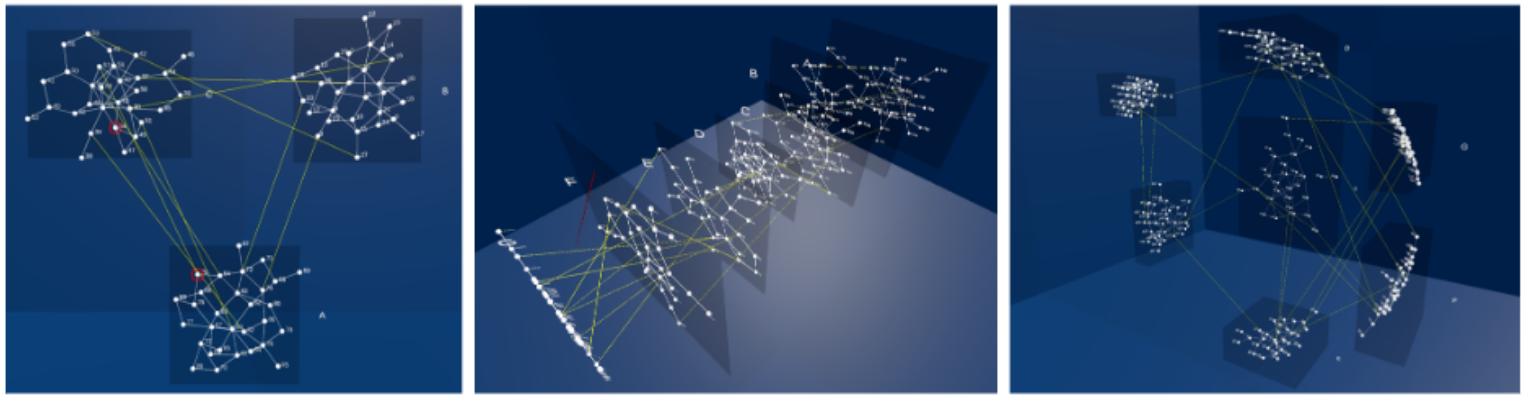
Source: Examples of visualisations in a virtual reality setting (2D, 2.5D, 3D) used during the experiment.



- Immersive analytics is an emerging field in VR to facilitate data analysis
- Eye tracking provides valuable insights into usability issues and interaction patterns

Understanding Visualisations in VR with Eye Tracking (Bachelor/Master)

Source: Examples of visualisations in a VR setting (2D, 2.5D, 3D) used during the experiment.



Example research questions:

1. Could we detect when users find visualisations are too complex by eye tracking?
2. Is there a way to automatically initiate a proper scale of visualisations?

Contact me!

yao.wang@vis.uni-stuttgart.de

Open Theses Topics

Mayar

Research Interests

- Privacy-preserving eye tracking
  - Adversarial attacks
  - Post-quantum cryptography

Research goal

Advance privacy-preserving eye tracking methods through provable (cryptographic) privacy guarantees.

Saliency Inference Attack (Master)

- Co-supervised by Yao Wang
- Goal: Perform a new side channel attack for model information stealing, i.e., infer information about the SalChartQA dataset.
- Requirements: Python (Pytorch/Tenserflow)

Contact me!

mayar.elfares@vis.uni-stuttgart.de

Open Theses Topics

Anna

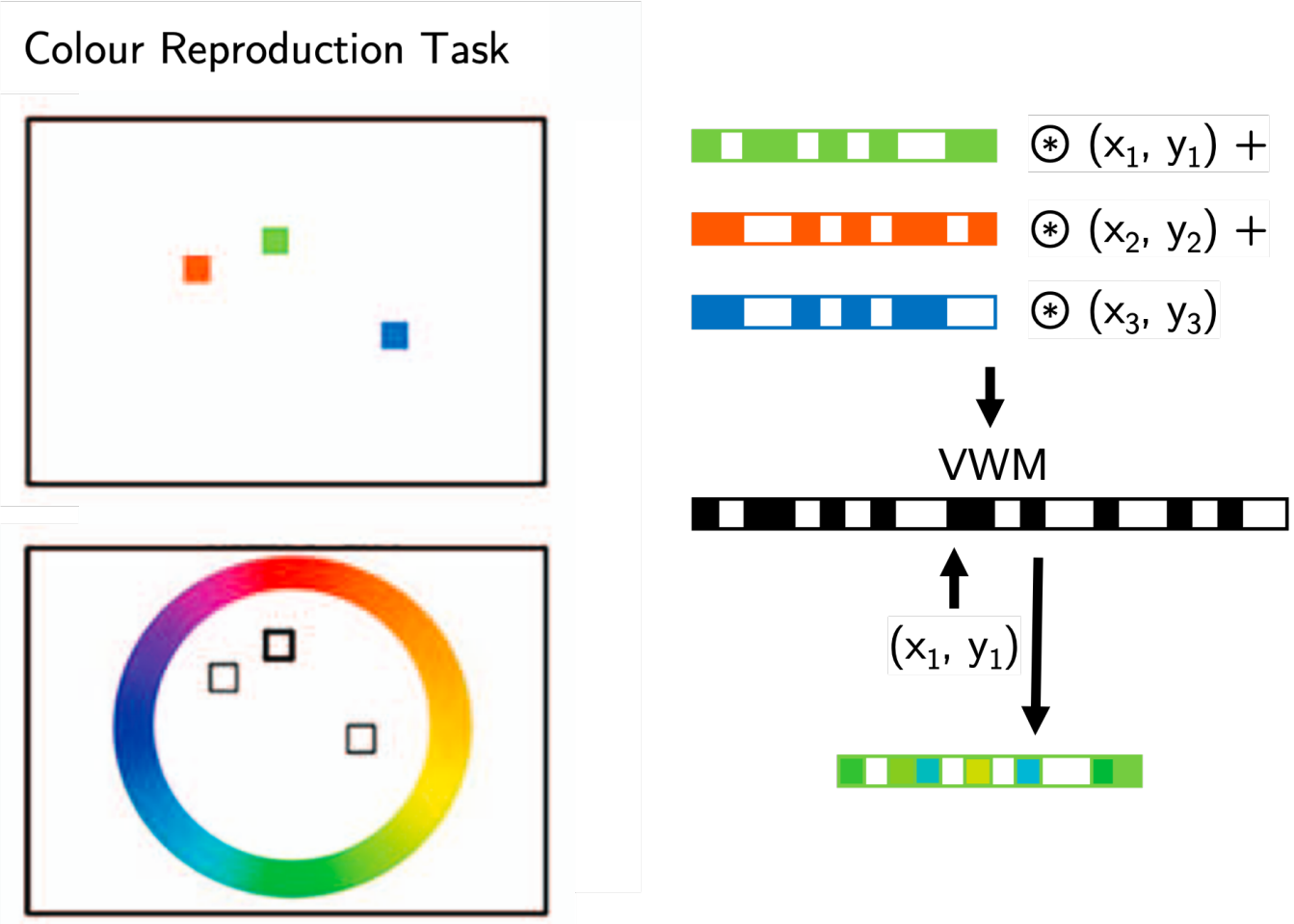
Cognitive Modeling Frameworks

- Rational Analysis (ACT-R) Anderson et al. [2004]
- Neural Engineering Framework (NEF) Eliasmith [2013]
- Computational Rationality Oulasvirta et al. [2022]

Research goal

Understanding and computationally **modelling human visual perception** for Human-Computer-Interaction (HCI) and Reinforcement Learning (RL).

Cognitively-plausible Mental Image Representation (Bachelor)



Source: Oberauer and Lin [2017]

Cognitively-plausible Mental Image Representation (Bachelor)

1. Implement an interface for the colour reproduction task
2. Implement a Visual Working Memory (VWM) model with Spatial Semantic Pointers (SSPs) Eliasmith [2013]
3. Compare VWM model to human data Oberauer and Lin [2017]
4. Master: extend interface for more complex tasks

10% literature review

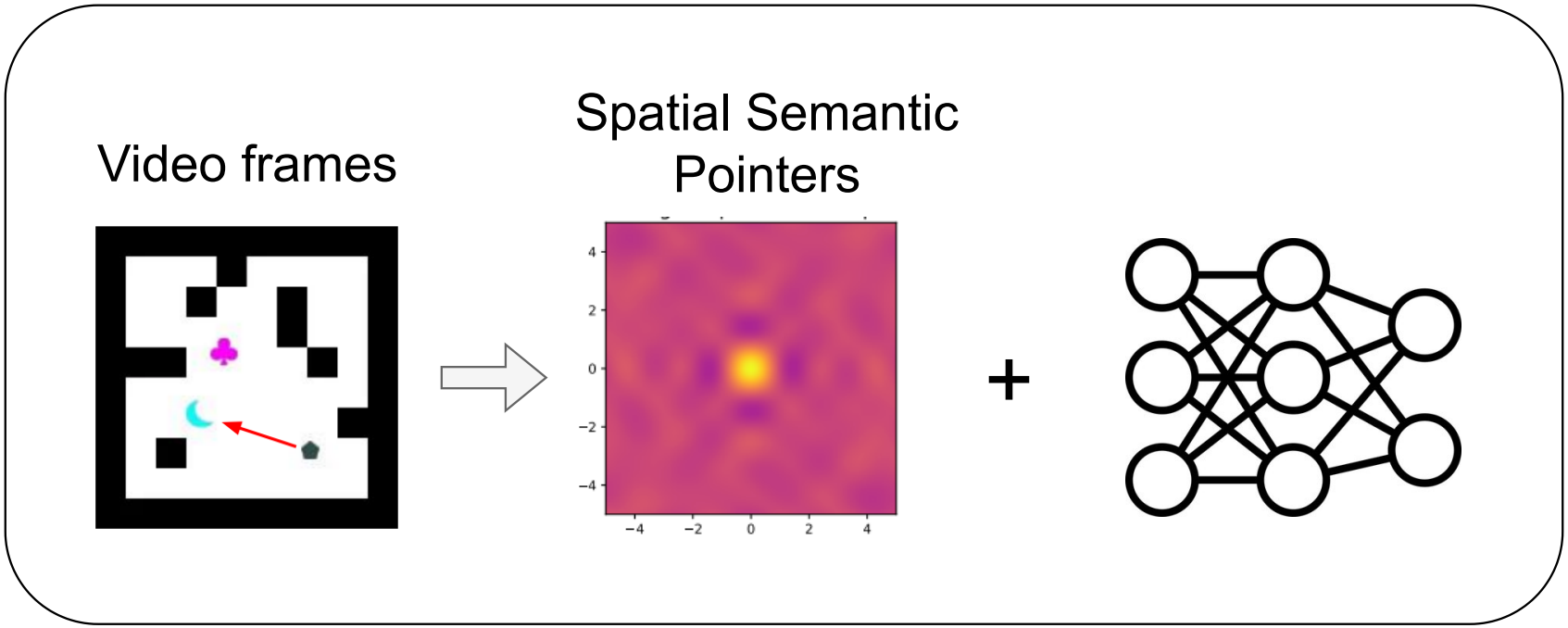
20% data processing

60% implementation

10% analysis

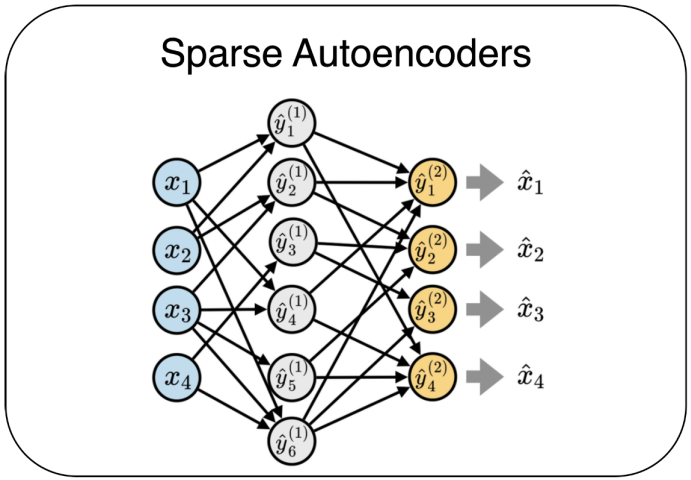
Neural Reasoning with cognitively-inspired Representations (Bachelor) – co-supervised with Matteo ♡

Goal: Use SSP Komer and Eliasmith [2020] to encode video frames and perform intuitive reasoning tasks Gandhi et al. [2021]; Bortoletto et al. [2024].

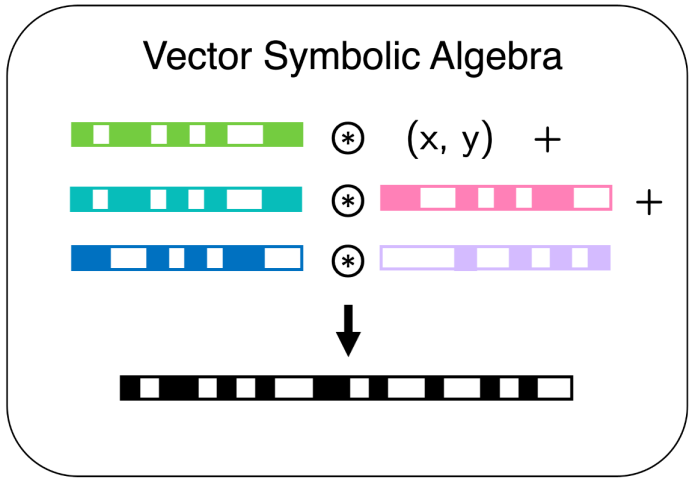


10% literature review | 70% implementation | 20% analysis

A Vector Symbolic Algebra for Language Models (Master) – co-supervised with Matteo ♡



meet



A Vector Symbolic Algebra for Language Models (Master) – co-supervised with Matteo ♡

1. Select a suitable NLP task e.g. IOI Wang et al. [2022]
   - *When Matteo and Anna went to the store, Matteo gave a drink to* _____
2. Analyze an LM (Pythia-70m) features obtained via Sparse Auto-Encoders Cunningham et al. [2023]
3. Explicitly model the task with a VSA Furlong and Eliasmith [2023]
4. Analyze and align both representations

20% literature research | 40% implementation | 40% analysis

Human Reasoning Module (Bachelor/ Master) – co-supervised with Ekta ♡

Problem

1. B is on the right of A
2. C is on the left of B
3. D is below C
4. E is below B
What is the relation between D and E?

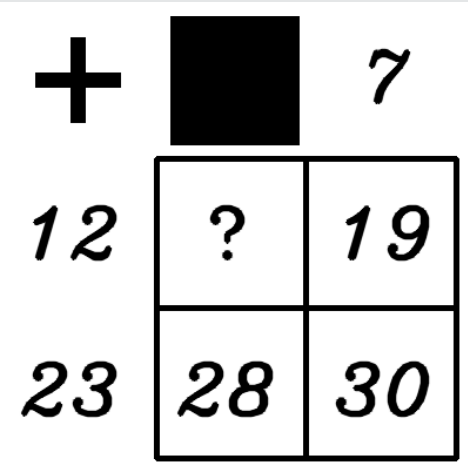
Possible mental states:

C A B
D E

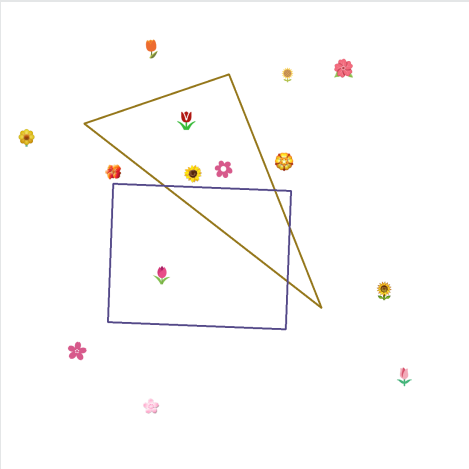
A C B
D E

Source: Nyamsuren and Taatgen [2014]

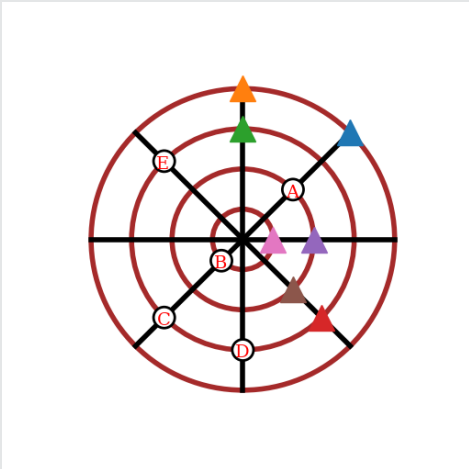
Algebra



Counting



Spatial Reasoning



Source: Cherian et al. [2023]

Human Reasoning Module (Bachelor/ Master) – co-supervised with Ekta ♡

1. Implement the Human Reasoning Module (HRM) in Python

   Nyamsuren and Taatgen [2014]

2. Simulate and verify data for simple reasoning tasks (causal deduction, spatial relations, inference of cause/effect)

3. Apply HRM to the visual question answering data set SMART101 Cherian et al. [2023]

20% literature research | 40% implementation | 40% analysis

Contact me!

anna.penzkofer@vis.uni-stuttgart.de

Open Theses Topics

Lei

Research Interests

- Eye tracking
- Intention prediction
- Graph neural networks
- Action recognition

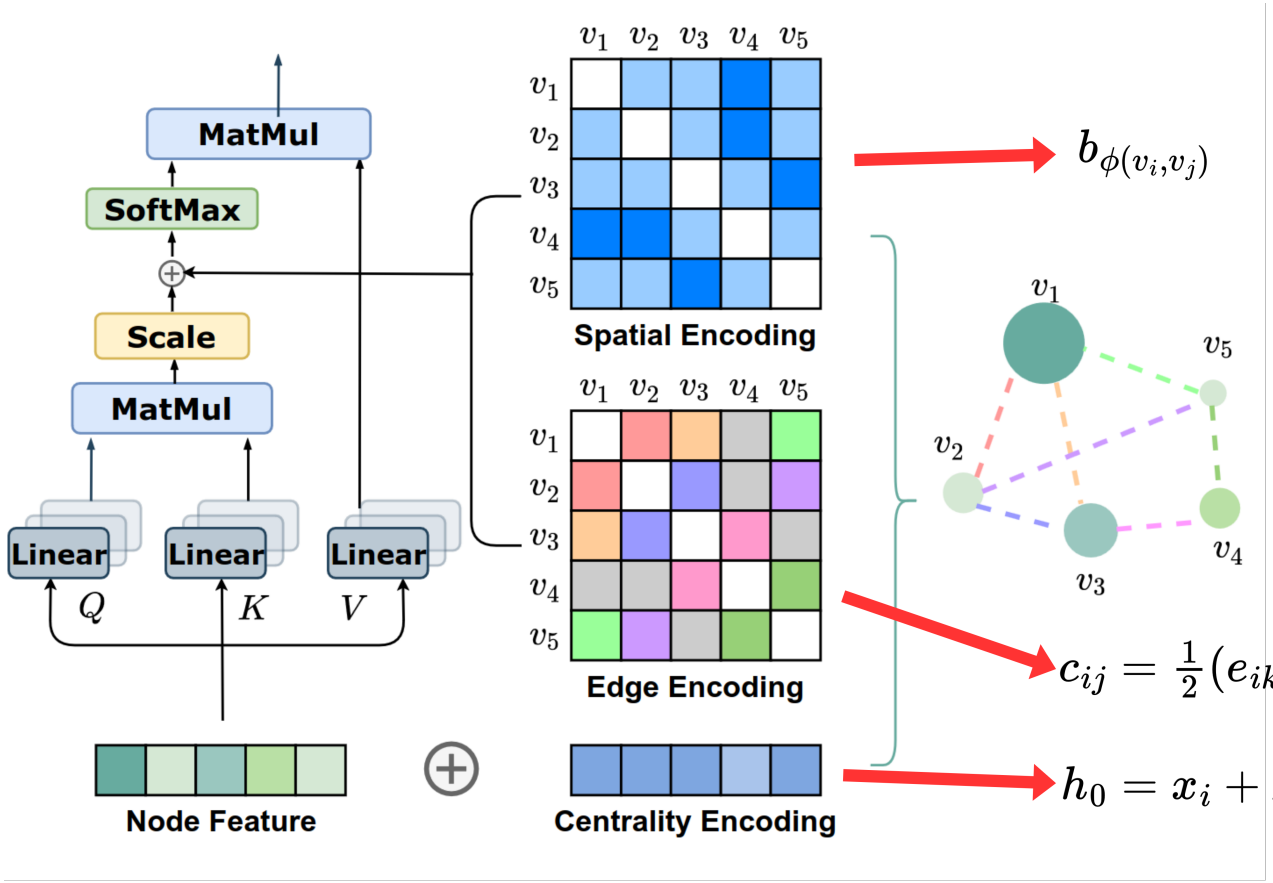
Research goal

Bayesian Intention Prediction in Human-AI Collaboration

Multi-modal Graphomer for Action Recognition in Egocentric Videos (Master)

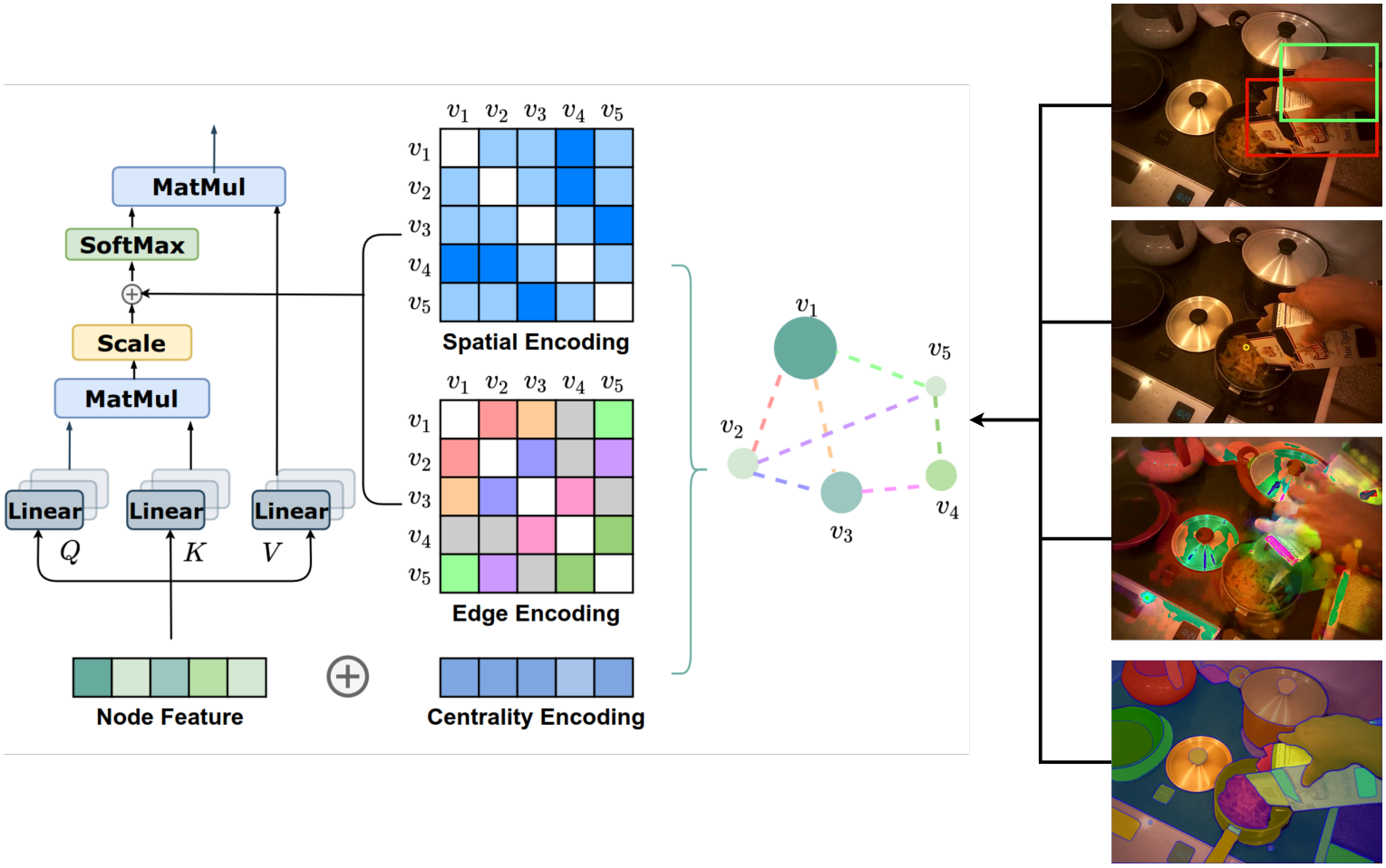
- Hand-object interaction
- Eye-hand coordination
- Motion objects, hand(s) and gaze
- Scene semantics

Multi-modal Graphomer for Action Recognition in Egocentric Videos (Master)



Source: Ying et al. [2021]

Multi-modal Graphomer for Action Recognition in Egocentric Videos (Master)



Source: Ying et al. [2021]

Contact me!

lei.shi@vis.uni-stuttgart.de

Table of Contents

Next Steps

Before the thesis

Before the thesis

- Contact the researcher to discuss the thesis topic
- Develop a project proposal with our supervision: motivation, problem statement/mandatory and optional goals, related work, approach/method, intended outcomes, schedule with milestones
  - Serves as a great basis for your written BSc/MSc thesis
- General requirements
  - Have very good/excellent grades
  - Ability to work independently and pro-actively

Working on the thesis

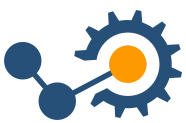
- Fully embedded in our group (workstation in lab, invitation to our group meetings, ...)
- Access to state-of-the-art hardware (GPU servers, eye trackers, on-body sensors, motion tracking, VR/AR headsets, ...)
- Weekly meetings of at least 30 minutes
- 3 presentations: Intro presentation, Mid-term presentation (graded), Final presentation (graded)

After the thesis

- Evaluation form
  - 40% General (commitment, methodology, autonomy, creativity, theory, general understanding)
  - 20% Implementation (appropriateness, documentation and reusability)
  - 20% Report (structure, accuracy, completeness)
  - 20% Presentations (motivation, structures, presentation materials, critical reflection, discussion skills, pace, presentations skills)
- Aim for a submission at conference/journal

Get in touch!

- Now - we have time for individual discussions
- Join the Ilias group for slides: Open Theses at Human-Computer Interaction and Cognitive Systems
- Find contact details on https://www.perceptualui.org/people/
- Find open projects throughout the year: https://www.perceptualui.org/teaching/open-projects/
- Join our lectures:
  - Practical Course Interactive Systems: Computational Theory of Mind and Cognition (summer term)
  - Medieninformatik (winter term)
  - Machine Perception and Learning (winter term)



Thank you!

Questions?

References i

J. R. Anderson, D. Bothell, M. D. Byrne, S. Douglass, C. Lebiere, and Y. Qin. An Integrated Theory of the Mind. *Psychological Review*, 111(4):1036–1060, 2004. ISSN 1939-1471. doi: 10.1037/0033-295X.111.4.1036.

C.-P. Bara, Z. Ma, Y. Yu, J. Shah, and J. Chai. Towards collaborative plan acquisition through theory of mind modeling in situated dialogue. *arXiv preprint arXiv:2305.11271*, 2023.

N. Bard, J. N. Foerster, S. Chandar, N. Burch, M. Lanctot, H. F. Song, E. Parisotto, V. Dumoulin, S. Moitra, E. Hughes, I. Dunning, S. Mourad, H. Larochelle, M. G. Bellemare, and M. Bowling. The hanabi challenge: A new frontier for ai research. *Artificial Intelligence*, 280:103216, 2020. ISSN 0004-3702. doi: https://doi.org/10.1016/j.artint.2019.103216. URL `https://www.sciencedirect.com/science/article/pii/S0004370219300116`.

M. Bortoletto, L. Shi, and A. Bulling. Neural reasoning about agents' goals, preferences, and actions. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, pages 456–464, 2024.

A. Chandrasekaran, D. Yadav, P. Chattopadhyay, V. Prabhu, and D. Parikh. It takes two to tango: Towards theory of ai's mind, 2017.

References ii

A. Cherian, K.-C. Peng, S. Lohit, K. A. Smith, and J. B. Tenenbaum. Are Deep Neural Networks SMARTer Than Second Graders? In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 10834–10844, Vancouver, BC, Canada, June 2023. IEEE. ISBN 9798350301298. doi: 10.1109/CVPR52729.2023.01043. URL https://ieeexplore.ieee.org/document/10204961/.

H. Cunningham, A. Ewart, L. Riggs, R. Huben, and L. Sharkey. Sparse Autoencoders Find Highly Interpretable Features in Language Models, Oct. 2023. URL http://arxiv.org/abs/2309.08600.

C. Eliasmith. *How to Build a Brain: A Neural Architecture for Biological Cognition*. Oxford University Press, June 2013. ISBN 978-0-19-979454-6. doi: 10.1093/acprof:oso/9780199794546.001.0001. URL https://academic.oup.com/book/6263.

P. M. Furlong and C. Eliasmith. Bridging Cognitive Architectures and Generative Models with Vector Symbolic Algebras. *Proceedings of the AAAI Symposium Series*, 2(1):262–271, 2023. doi: 10.1609/aaaiss.v2i1.27686. URL https://ojs.aaai.org/index.php/AAAI-SS/article/view/27686.

References iii

K. Gandhi, G. Stojnic, B. M. Lake, and M. Dillon. Baby intuitions benchmark (BIB): Discerning the goals, preferences, and actions of others. In *Thirty-Fifth Conference on Neural Information Processing Systems*, 2021. URL `https://arxiv.org/abs/2102.11938`.

D. Hafner, T. Lillicrap, J. Ba, and M. Norouzi. Dream to control: Learning behaviors by latent imagination. In *International Conference on Learning Representations*, 2020.

D. Hafner, T. P. Lillicrap, M. Norouzi, and J. Ba. Mastering atari with discrete world models. In *International Conference on Learning Representations*, 2021.

D. Han, J. Choe, S. Chun, J. J. Y. Chung, M. Chang, S. Yun, J. Y. Song, and S. J. Oh. Neglected free lunch – learning image classifiers using annotation byproducts. In *International Conference on Computer Vision (ICCV)*, 2023.

References iv

H. He, J. Boyd-Graber, K. Kwok, and H. Daumé, III. Opponent modeling in deep reinforcement learning. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 1804–1813, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `https://proceedings.mlr.press/v48/he16.html`.

B. Komer and C. Eliasmith. Efficient navigation using a scalable, biologically inspired spatial representation. 2020. URL `https://www.semanticscholar.org/paper/Efficient-navigation-using-a-scalable%2C-biologically-Komer-Eliasmith/fa49a5a04cb8b2bbfb7954cdb2c45e8bf7f8a5ff`.

Latoschik VRST '17. The effect of avatar realism in immersive social virtual realities. In *Proceedings of the 2017 ACM Symposium on Virtual Reality Software and Technology*, pages 1–10, 2017.

S. Marks and M. Tegmark. The geometry of truth: Emergent linear structure in large language model representations of true/false datasets, 2023.

Meta. Meta horizon worlds. `https://www.meta.com/horizon-worlds/`.

E. Nyamsuren and N. A. Taatgen. Human Reasoning Module. *Biologically Inspired Cognitive Architectures*, 8:1–18, Apr. 2014. doi: 10.1016/j.bica.2014.02.002. URL https://www.sciencedirect.com/science/article/pii/S2212683X14000097.

K. Oberauer and H.-Y. Lin. An interference model of visual working memory. *Psychological Review*, 124(1):21–59, 2017. doi: 10.1037/rev0000044.

A. Oulasvirta, J. P. P. Jokinen, and A. Howes. Computational Rationality as a Theory of Interaction. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, pages 1–14, New York, NY, USA, Apr. 2022. Association for Computing Machinery. ISBN 978-1-4503-9157-3. doi: 10.1145/3491102.3517739. URL https://doi.org/10.1145/3491102.3517739.

N. Rabinowitz, F. Perbet, F. Song, C. Zhang, S. M. A. Eslami, and M. Botvinick. Machine theory of mind. In J. Dy and A. Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 4218–4227. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/rabinowitz18a.html.

References vi

R. Raileanu, E. L. Denton, A. Szlam, and R. Fergus. Modeling others using oneself in multi-agent reinforcement learning. In *International Conference on Machine Learning*, 2018. URL https://api.semanticscholar.org/CorpusID:3622509.

D. D. Salvucci. An integrated model of eye movements and visual encoding. *Cognitive Systems Research*, 1(4):201–220, Feb. 2001. ISSN 1389-0417. doi: 10.1016/S1389-0417(00)00015-2. URL https://www.sciencedirect.com/science/article/pii/S1389041700000152.

M. R. Samsami, A. Zholus, J. Rajendran, and S. Chandar. Mastering memory tasks with world models. In *The Twelfth International Conference on Learning Representations*, 2024.

A. Saran, R. Zhang, E. S. Short, and S. Niekum. Efficiently Guiding Imitation Learning Agents with Human Gaze, Apr. 2021. URL http://arxiv.org/abs/2002.12500.

M. Sclar, G. Neubig, and Y. Bisk. Symmetric machine theory of mind. In *International Conference on Machine Learning*, pages 19450–19466. PMLR, 2022.

K. Wang, A. Variengien, A. Conmy, B. Shlegeris, and J. Steinhardt. Interpretability in the Wild: a Circuit for Indirect Object Identification in GPT-2 small, Nov. 2022. URL http://arxiv.org/abs/2211.00593.

Y. Wang, W. Wang, A. Abdelhafez, M. Elfares, Z. Hu, M. Bâce, and A. Bulling. Salchartqa: Question-driven saliency on information visualisations. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems (CHI)*, pages 1–14, 2024. doi: 10.1145/3613904.3642942.

C. Ying, T. Cai, S. Luo, S. Zheng, G. Ke, D. He, Y. Shen, and T.-Y. Liu. Do transformers really perform badly for graph representation? *Advances in neural information processing systems*, 34:28877–28888, 2021.

X. Yu, J. Jiang, W. Zhang, H. Jiang, and Z. Lu. Model-based opponent modeling. In *Proceedings of the 36th International Conference on Neural Information Processing Systems*, NIPS '22, Red Hook, NY, USA, 2022. Curran Associates Inc. ISBN 9781713871088.

R. Zhang, C. Walshe, Z. Liu, L. Guan, K. Muller, J. Whritner, L. Zhang, M. Hayhoe, and D. Ballard. Atari-HEAD: Atari Human Eye-Tracking and Demonstration Dataset. *Proceedings of the AAAI Conference on Artificial Intelligence*, 34(04): 6811–6820, Apr. 2020. ISSN 2374-3468. doi: 10.1609/aaai.v34i04.6161. URL `https://ojs.aaai.org/index.php/AAAI/article/view/6161`.

W. Zhu, Z. Zhang, and Y. Wang. Language models represent beliefs of self and others. *arXiv preprint arXiv:2402.18496*, 2024.