



Smart Tools and Applications in Graphics

November 27th-28th, 2025, Genoa, Italy

GUIDÆTA

A Versatile Interactions Dataset
with extensive Context Information and Metadata

**S. Lengauer¹, S.A. von Götz², M.T. Hoesch², F. Steinwider¹, M. Tytarenko¹, M.A. Bedek²,
and T. Schreck¹**

¹Graz University of Technology, Institute of Visual Computing, Austria

²University of Graz, Department of Psychology, Austria



Interaction Datasets

Data

- Mouse/(keyboard) events: timestamp, position, ...
- (Metadata)
- (Eye Tracking)

Application domains

- Cognitive sciences
- Human computer interaction → *interaction ergonomics*
- (Adaptive) Visualization
- Cybersecurity → *behavioral biometrics*

Interaction Datasets

Strengths of the GUID/ETA dataset

- 1 *Guided* interactions based on 3 information retrieval tasks
- 2 Context + widget information for events
- 3 Rich metadata (sociodemographic information, explicit feedback, cognitive load)
- 4 Data and tooling publicly available, respecting the FAIR^[12] principles

Large in scale → good balance between # of participants & individual interaction duration.

^[12]M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1 (2016), pp. 1–9

Interaction Datasets

Strengths of the GUID/ETA dataset

- 1 *Guided* interactions based on 3 information retrieval tasks
- 2 Context + widget information for events
- 3 Rich metadata (sociodemographic information, explicit feedback, cognitive load)
- 4 Data and tooling publicly available, respecting the FAIR^[12] principles

Large in scale → good balance between # of participants & individual interaction duration.

^[12]M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1 (2016), pp. 1–9

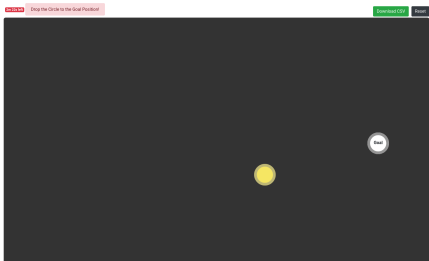
Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

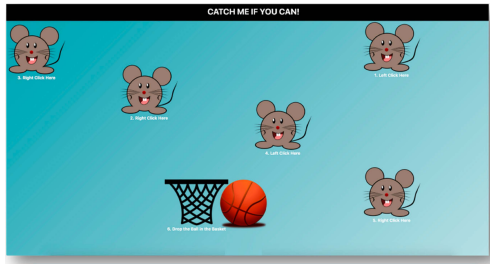
Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

Related Interaction Datasets



SapiMouse^[3]



ReMouse^[10]

^[3]M. Antal, N. Fejér, and K. Buza. “SapiMouse: Mouse Dynamics-based User Authentication Using Deep Feature Learning”. In: *2021 IEEE 15th International Symposium on (SACI)*. May 2021, pp. 61–66. DOI: 10.1109/SACI51354.2021.9465583

^[10]S. Sadehpour and N. Vljic. “ReMouse Dataset: On the Efficacy of Measuring the Similarity of Human-Generated Trajectories for the Detection of Session-Replay Bots”. In: *Journal of Cybersecurity and Privacy 3.1* (2023), pp. 95–117. ISSN: 2624-800X. DOI: 10.3390/jcp3010007

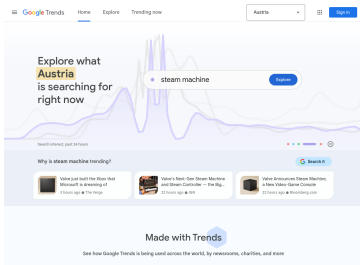
Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

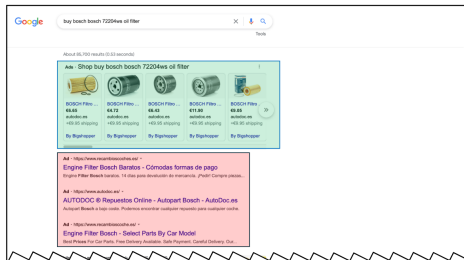
Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

Related Interaction Datasets



The Attentive Cursor^[8]



AdSERP^[7]

[8] L. A. Leiva and I. Arapakis. "The Attentive Cursor Dataset". In: *Frontiers in Human Neuroscience Volume 14 - 2020 (2020)*. ISSN: 1662-5161. DOI: 10.3389/fnhum.2020.565664

[7] K. Latifzadeh, J. Gwizdka, and L. A. Leiva. "A Versatile Dataset of Mouse and Eye Movements on Search Engine Results Pages". In: *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2025, pp. 3412–3421.

Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

Related Interaction Datasets

Dataset	# Users	# Interactions	Observation duration	Widget Information	Guided
Balabit ^[4]	10	4.61 M	178.21 H		
Chao Shen ^[11]	28	79.69 M	499.00 H		
Bogazici ^[5]	24	72.83 M	539.62 H	✓	
DFL ^[2]	21	129.56 M	718.06 H		
SapiMouse ^[3, 1]	120	1.18 M	8.22 H	N.A.	✓
ReMouse ^[10]	100	0.38 M	6.74 H	N.A.	✓
The Attentive Cursor ^[8]	2,909	0.11 M	12.74 H	✓	(✓)
AdSERP ^[7]	47	1.17 M	16.39 H		(✓)
GUID/ETA	253	2.39 M	49.61 H	✓	✓

IVC

The A⁺CHIS Project

A+CHIS Erkennung Lesezeichen Aufgaben

Willkommen, IDK55Rh Abmelden

2.4 Wie wollen Sie leben? - Ihre Entscheidungen

Ihr Arzt verfügt über das medizinische Wissen, um Sie wirksam zu behandeln. Doch allein erreicht er nicht viel. Denn im Mittelpunkt der Behandlung stehen Sie - Ihre Einstellung ist entscheidend. Das ist ganz wörtlich gemeint: In der Behandlung und in Ihrem Alltag werden Sie tatsächlich immer wieder Entscheidungen treffen. Denn nichts passiert ohne Ihre Zustimmung. Sogar wenn Sie alles dem Arzt überlassen und sich selbst möglichst wenig mit Diabetes befassen, treffen Sie eine Entscheidung - gegen Ihre aktive Mitarbeit. Viele Patienten profitieren allerdings davon, wenn Sie die Therapie mitgestalten und sich eigene Ziele setzen - zum Beispiel, was den HbA1c-Wert oder das Gewicht angeht.

2.4.1 Gesund leben heißt, viele kleine Entscheidungen zu treffen

Überlegen Sie gemeinsam mit dem Arzt möglichst konkrete Ziele und wie Sie diese erreichen können. Dabei geht es oft gar nicht um rein medizinische Fragen. Mindestens genauso wichtig sind die praktischen Entscheidungen im Alltag: Gibt es abends Offensmüse oder Schweinracken? Nehmen Sie die Treppe oder den Aufzug? Halten Sie sich zum geplanten Gang durch den Park auf, obwohl es gerade anfangt zu regnen? Die Summe solcher Entscheidungen macht Ihr Gesundheitsverhalten aus - und das hat großen Einfluss auf Ihren Diabetes.

Vorkommen: Insulin

Men Körper

Blutzucker Mittel

Ursache Insulin

Depression Dosis

Problem

Zucker mmol

Menge Arzt

Behandlung

T1 T2 T3 T4 T5

5 Unterzuckerungen rechtzeitig erkennen - die Symptome

5.1.1 Unterzuckerungen rechtzeitig erkennen - die Symptome

[...] Deshalb reagiert der **Körper**, wenn der Blutzucker sinkt, wie auf eine Kampfstation: Er schüttet das Stresshormon Adrenalin aus. [...]

5.1.2 Unterzuckerungen wirksam behandeln - Gegenmaßnahmen

[...] Die Regel lautet: „Bist essen, dann messen.“ Das Wichtigste ist, dem **Körper** sofort Kohlenhydrate zuzuführen, die direkt ins Blut gelangen. Stoppen Sie außerdem sofort **Körper**-liche Aktivitäten - sie senken den Blutzucker weiter. [...]

[...] Falls nicht, brauche der **Körper** mehr schnelle Kohlenhydrate (Kap. 3.2.2).

5.1.3 Wie kommt es zur Unterzuckerung? - Ursachen

[...] Dort tragen Sie Ihre Blutzuckerwerte ein, vermerken, wann Sie welche Medikamente nehmen, aber auch Mahlzeiten, Getränke, **Körper**-liche Anstrengungen und eventuelle Beschwerden. [...]

[...] Grundsätzlich besteht die Gefahr einer Hypoglykämie immer dann, wenn die Menge der gegessenen Kohlenhydrate, die Dosis der Medikamente oder des gesparten Insulins, **Körper**-liche Anstrengung, Alkoholkonsum oder andere zucker senkende Einflüsse nicht aufeinander abgestimmt sind. [...]

[...] durch Spritzen in einen Muskel oder weil Hitze die Durchblutung stark anregt: 4) Ihr Insulinbedarf ist gesunken, weil Sie abgenommen haben: nehmen Sie Sulfonharnstoffe oder Glime, sind mögliche Ursachen: 1) Sie haben aus Versehen eine Dosis doppelt eingenommen 2) Sie haben nach der Einnahme nicht gegessen 3) Sie haben abgenommen und brauchen eine geringere Dosis 4) Wechselwirkung mit anderen Medikamenten

- Sport und **Körper**-liche Anstrengung: Regelmäßige Bewegung senkt langfristig den Blutzucker. [...]
- Infekte: Vor allem bei Magen-Darm-Infekten verliert der **Körper** auch Kohlenhydrate. Dann ist eventuell die gespritzte Insulinmenge oder die Dosis der Tabletten zu hoch.

Detaillierte Suche

Suchbegriff: **Körper** Vorkommen: 44 (in gesamten Dokumenten)

Diagramm 1: „Ich habe Diabetes“ - was bedeutet das?

Diagramm 2: Auf gutem Weg - wie viele Diabetes



Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

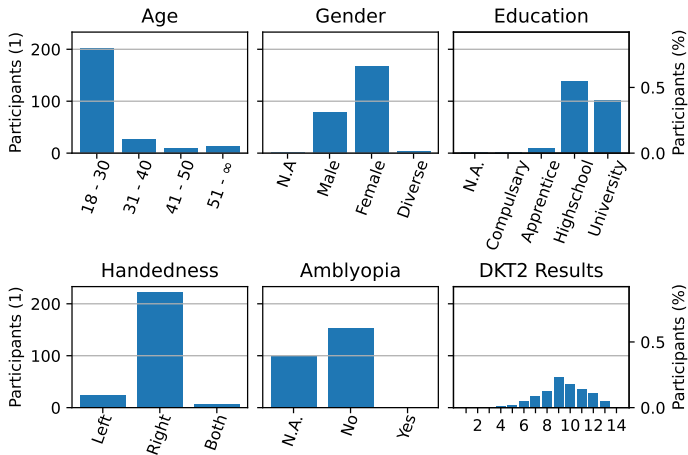
Data Collection

Cognitive Science Study (May 12 – June 23, 2025)

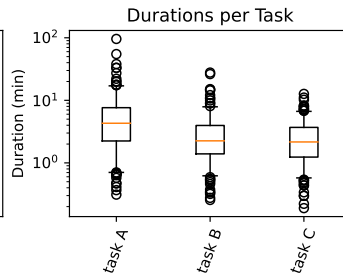
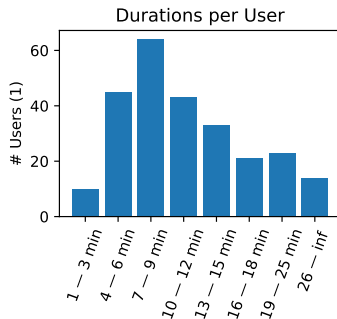
- Diverse advertisement channels
- Reimbursement of €20 upon completion + 10× additional €100
- Completely unsupervised: participants receive link and work through 5 steps

- 1 Welcoming / Study Overview Information / GDPR Agreement
- 2 Pre-tests: Sociodemographics, Pre-knowledge
- 3 Onboarding through help wizard
- 4 Solving of 3 **randomized** IR tasks
- 5 SUS and sendoff

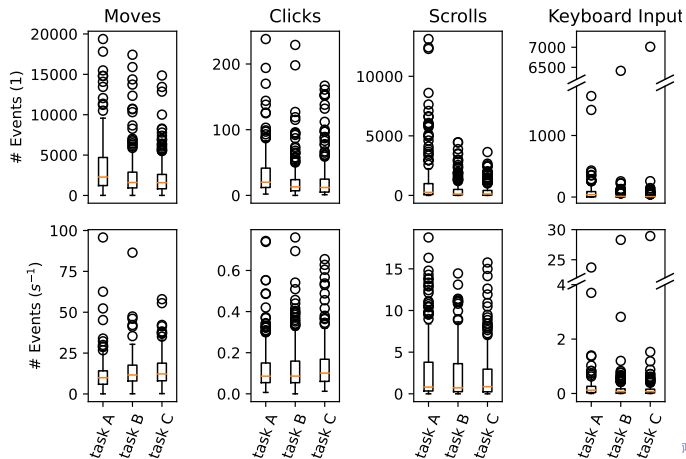
Sociodemographics



Task-related Statistics

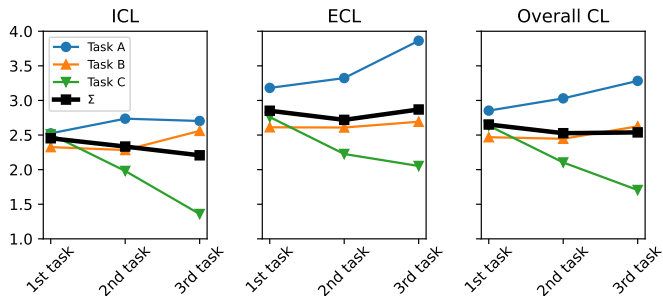


Task-related Statistics



Cognitive Load

Combination of Intrinsic (ICL) and Extrinsic Cognitive Load (ECL)^[6]



^[6]F. Krieglstein, M. Beege, G. D. Rey, C. Sanchez-Stockhammer, and S. Schneider. "Development and validation of a theory-based questionnaire to measure different types of cognitive load". en. In: *Educational Psychology Review* 35.1 (2023), p. 9. ISSN: 1040-726X, 1573-336X. DOI: 10.1007/s10648-023-09738-0

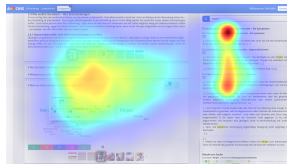
Cursor Positions



task A



task B



task C



Combined

InputEvents

mouse_event

id	integer
session_id	integer NN
timestamp	timestamp
type	enum
client_x	integer
client_y	integer
chapter	integer
component	enumerate
context	json

keyboard_event

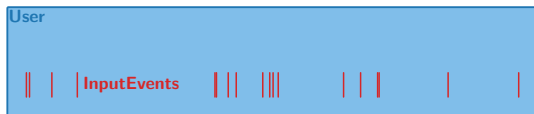
id	integer
session_id	integer NN
timestamp	timestamp
key	char
modifiers	integer
context	json

Event Containers

|| | | **InputEvents** ||| |||| | | | |

———— Time —→

Event Containers



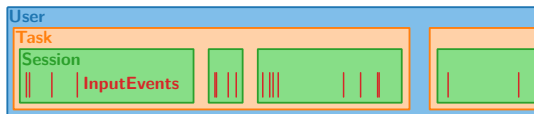
user	
id	integer
created_at	timestamp
age	integer
gender	enum
education	enum
handedness	enum
amblyopia	enum
[dkt2answers]	char
[dkt2aswer_durations]	integer
[sus_scales]	integer
activity	float

Event Containers



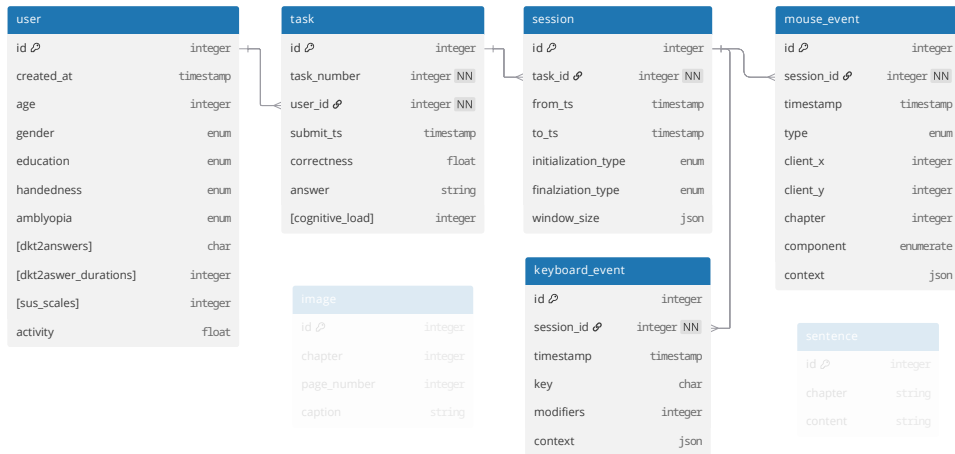
task	
id	integer
task_number	integer NN
user_id	integer NN
submit_ts	timestamp
correctness	float
answer	string
[cognitive_load]	integer

Event Containers

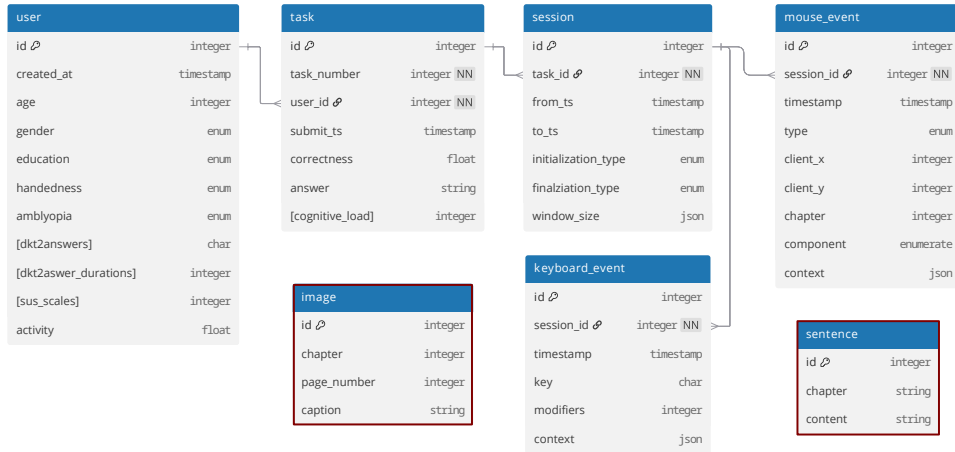


session	
id	integer
task_id	integer NN
from_ts	timestamp
to_ts	timestamp
initialization_type	enum
finalization_type	enum
window_size	json

Schema



Schema



Schema



Data Management

The screenshot shows the GitHub repository page for 'apchis-guidaeta' by 'lengauer'. The repository is public and has 0 stars, 0 forks, and 0 watchers. The main branch is 'main'. The repository contains a README.md file, a LICENSE file, and a main.py file. The README.md file is selected and shows the title 'GUIDÆTA' and a description: 'A Versatile Interactions Dataset with extensive Context Information and Metadata'. The description mentions that the dataset was collected within an online user study in the A+CHIS project on Human-Centered Interactive Adaptive Visual Approaches in High-Quality Health Information, conducted between May 12 and June 23, 2025. It also mentions that extensive information on scope, data collection, metadata, preliminary analysis can be found in the respective paper 'GUIDÆTA - A Versatile Interactions Dataset with extensive Context Information and Metadata', published at Smart Tools and Applications in Graphics (STAG) 2025, an annual international conference organized by the Italian Chapter of the Eurographics association. The raw dataset is publicly available on OSP: <https://osf.io/fhvbmv/>.

The repository also has a 'Releases' section with no releases published and a 'Packages' section with no packages published. The 'Languages' section shows Python at 100.0%.



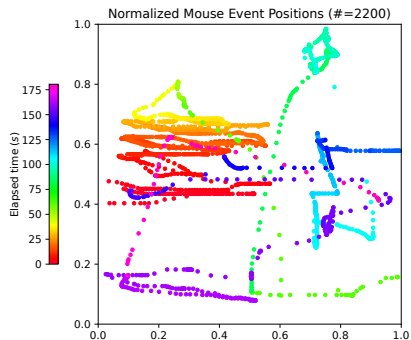
Data Management

Python framework for data handling and analysis

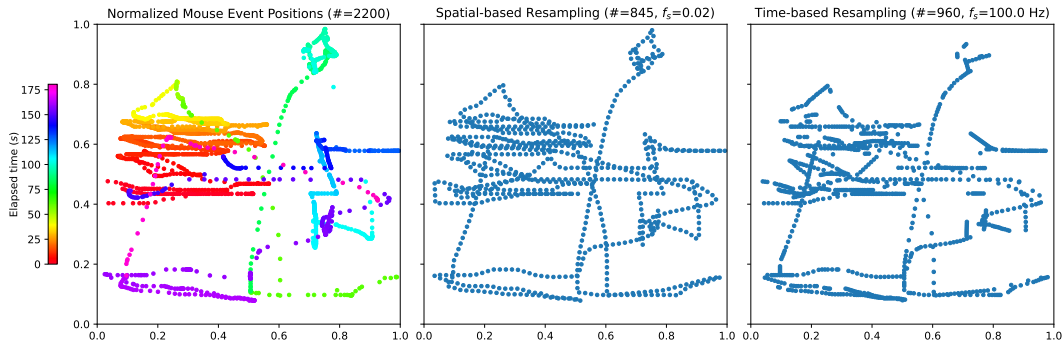
```
1 import observations
2
3 # load everything from files and initialize
4 sentences, sessions, users, task_answers = observations.load_data()
5
6 # get the cognitive load experienced for task two
7 task_two_answers = [ta for ta in task_answers if ta.task_no==2]
8 cl = sum([
9     sum(ta.cl.scores)/len(ta.cl.scores) for ta in task_two_answers
10 ])/len(task_two_answers)
11
12 # get all sessions associated with the 3rd task
13 s = [s for ta in task_answers for s in ta.sessions if ta.task_no==3]
```



Sampling

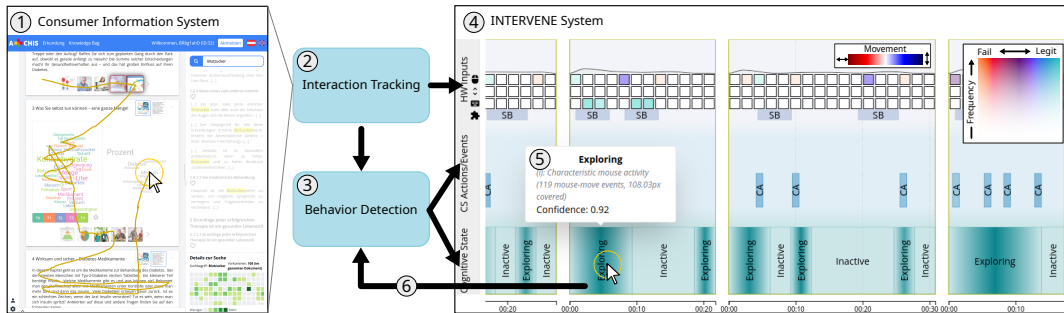


Sampling



Outlook

INTERVENE^[9]



^[9]S. Lengauer, L. Shao, M. Tytarenko, M. Klaffenböck, and T. Schreck. "Interaction Visualization for Analysing and Improving User Models".

In: 32nd ACM Conference on User Modeling, Adaptation and Personalization June 2024: UMAP'2024. 2024

Conclusion

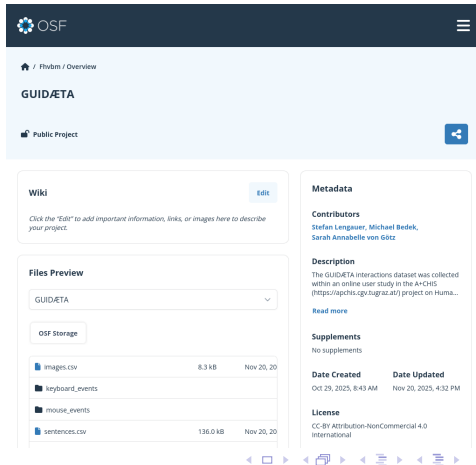
- Versatile interaction dataset obtained from 253 individual users
- Rich context and content information
- Tools for data handling and analysis



Dataset



Tooling



References I

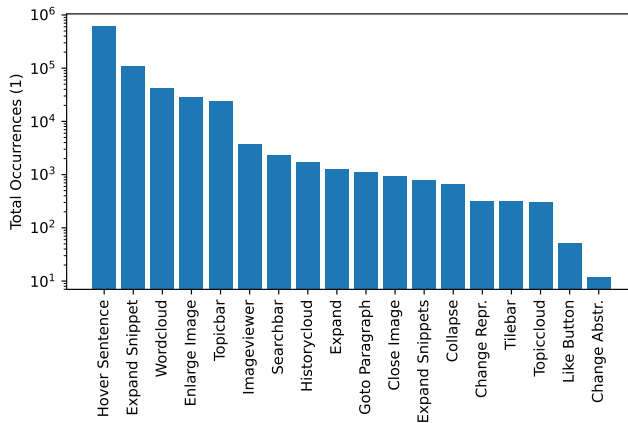
- [1] M. Antal, K. Buza, and N. Fejer. "SapiAgent: A Bot Based on Deep Learning to Generate Human-Like Mouse Trajectories". In: *IEEE Access* 9 (2021), pp. 124396–124408. DOI: 10.1109/ACCESS.2021.3111098. (Visited on 08/13/2025).
- [2] M. Antal and L. Denes-Fazakas. "User Verification Based on Mouse Dynamics: a Comparison of Public Data Sets". In: *2019 IEEE 13th International Symposium on Applied Computational Intelligence and Informatics (SACI)*. May 2019, pp. 143–148. DOI: 10.1109/SACI46893.2019.9111596. (Visited on 08/13/2025).
- [3] M. Antal, N. Fejér, and K. Buza. "SapiMouse: Mouse Dynamics-based User Authentication Using Deep Feature Learning". In: *2021 IEEE 15th International Symposium on (SACI)*. May 2021, pp. 61–66. DOI: 10.1109/SACI51354.2021.9465583.
- [4] Á. Fülöp, L. Kovács, T. Kurics, and E. Windhager-Pokol. *Balabit Mouse Challenge Data Set*. 2016.
- [5] A. A. Kılıç, M. Yıldırım, and E. Anarım. "Bogazici mouse dynamics dataset". In: *Data in Brief* 36 (June 2021), p. 107094. DOI: 10.1016/j.dib.2021.107094.
- [6] F. Krieglstein, M. Beege, G. D. Rey, C. Sanchez-Stockhammer, and S. Schneider. "Development and validation of a theory-based questionnaire to measure different types of cognitive load". en. In: *Educational Psychology Review* 35.1 (2023), p. 9. DOI: 10.1007/s10648-023-09738-0.
- [7] K. Latifzadeh, J. Gwizdka, and L. A. Leiva. "A Versatile Dataset of Mouse and Eye Movements on Search Engine Results Pages". In: *Proceedings of the 48th International ACM SIGIR Conference on Research and Development in Information Retrieval*. 2025, pp. 3412–3421.
- [8] L. A. Leiva and I. Arapakis. "The Attentive Cursor Dataset". In: *Frontiers in Human Neuroscience* Volume 14 - 2020 (2020). DOI: 10.3389/fnhum.2020.565664.

References II

- [9] S. Lengauer, L. Shao, M. Tytarenko, M. Klaffenböck, and T. Schreck. "Interaction Visualization for Analysing and Improving User Models". In: *32nd ACM Conference on User Modeling, Adaptation and Personalization June 2024: UMAP'2024*. 2024.
- [10] S. Sadeghpour and N. Vlajic. "ReMouse Dataset: On the Efficacy of Measuring the Similarity of Human-Generated Trajectories for the Detection of Session-Replay Bots". In: *Journal of Cybersecurity and Privacy* 3.1 (2023), pp. 95–117. DOI: 10.3390/jcp3010007.
- [11] C. Shen, Z. Cai, and X. Guan. "Continuous authentication for mouse dynamics: A pattern-growth approach". In: *IEEE/IFIP International Conference on Dependable Systems and Networks (DSN 2012)*. ISSN: 2158-3927. June 2012, pp. 1–12. DOI: 10.1109/DSN.2012.6263955. (Visited on 08/13/2025).
- [12] M. D. Wilkinson, M. Dumontier, I. J. Aalbersberg, et al. "The FAIR Guiding Principles for scientific data management and stewardship". In: *Scientific data* 3.1 (2016), pp. 1–9.

Context Information

35.13 % of **MouseEvents** feature context information.



File Structure

```
/
├── mouse_events
│   └── me_<user>_<task>.csv
├── keyboard_events
│   └── ke_<user>_<task>.csv
├── task_answers.csv
├── sessions.csv
├── users.csv
├── sentences.csv
└── images.csv
```

- All data given as CSV files, complying with FAIR^[12] principles
- Files (mostly) adhering to schema
- 1,316 Files
- 106 MB on disk