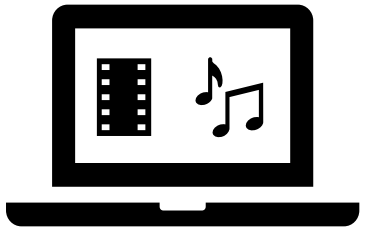
Data Awareness in Theory and Practice

Understanding Data Practices in Everyday Digital Technologies

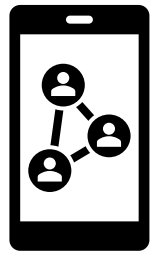
What is data awareness about?

Streaming services



Search engine

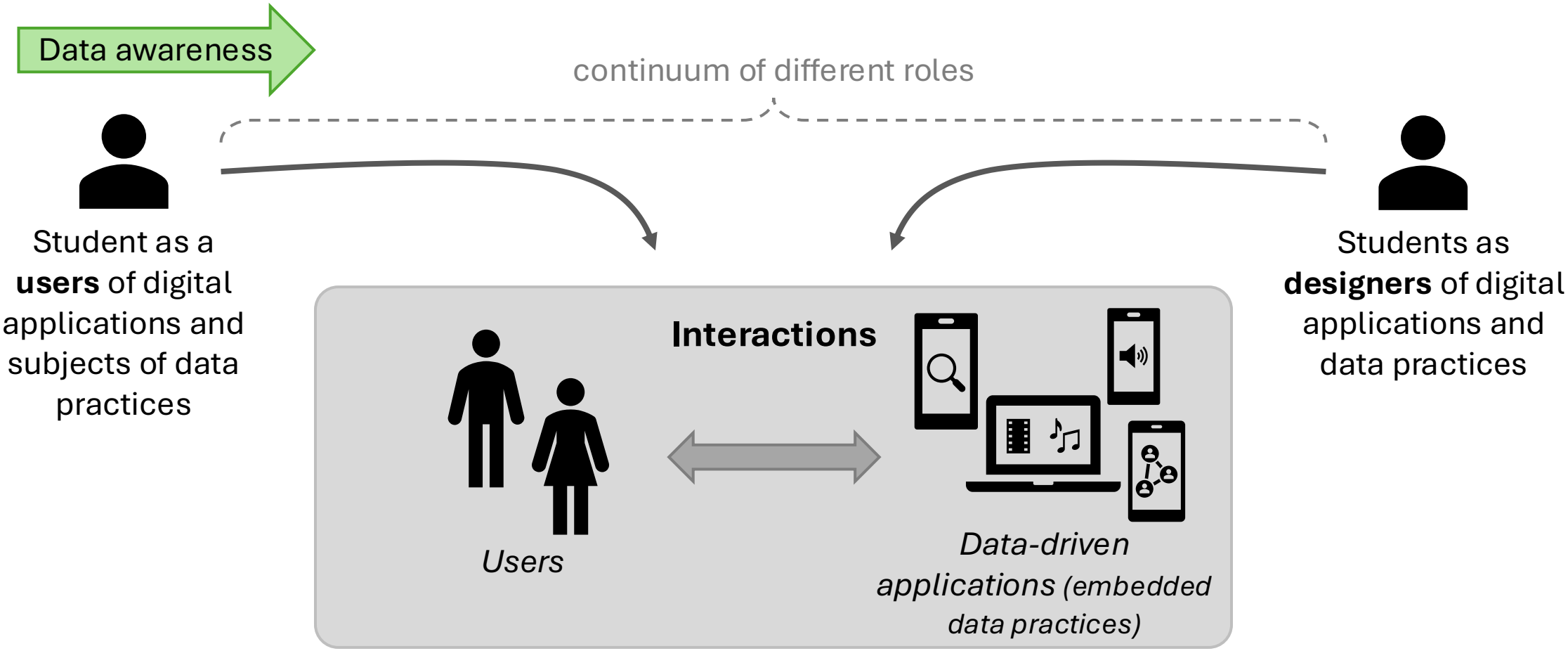
Social media



Data-driven technologies:

- Output and behavior strongly depends on collected data and the use of data-driven methods
- *In short*: ***what they do depends on what has happend earlier in the interaction process***

Starting point for data awareness



What is data awareness about?

1. Knowledge and skills to understand and reflect on the role of data in applications
2. Knowledge and skills learned in class should be meaningful and applicable in such situations

Challenges for these goals

1. Knowledge and skills to understand and reflect on the role of data in applications
2. Knowledge and skills learned in class should be meaningful and applicable in such situations

(1) Studies report a lacking understanding

See review in: Höper, L., & Schulte, C. (2024). The data awareness framework as part of data literacies in K-12 education. Information and Learning Sciences, 125(7/8), 491–512.

(2) What should they learn? Discussions about data and AI concepts

later more about that: explanatory model

Challenges for these goals

1. Knowledge and skills to understand and reflect on the role of data in applications
2. Knowledge and skills learned in class should be meaningful and applicable in such situations

(3) Students often struggle to relate learned concepts about data or ML to their daily lives

For example:

Vartiainen, H., et al. (2021). Machine learning for middle schoolers: Learning through data-driven design. International Journal of Child-Computer Interaction, 29, 1–12.

Bilstrup, K.-E. K., et al. (2022). Supporting Critical Data Literacy in K-9 Education: Three Principles for Enriching Pupils' Relationship to Data. Proceedings of the 21st Annual ACM Interaction Design and Children Conference, 225–236.

Challenges for these goals

1. Knowledge and skills to understand and reflect on the role of data in applications
2. Knowledge and skills learned in class should be meaningful and applicable in such situations

(4) Many students are not curious to understand the data practices in daily applications

See study with young school students:

Höper, L., Schulte, C., & Mühling, A. (2024). Students' Motivation and Intention to Engage with Data-Driven Technologies from a CS Perspective in Everyday Life. Proceedings of the 2024 Innovation and Technology in Computer Science Education V. 1 (ITiCSE 2024), 7.

Challenges for these goals

1. Knowledge and skills to understand and reflect on the role of data in applications

2. Knowledge and skills learned in class should be meaningful and applicable in such situations

(5) Privacy research indicates that many people just want to use digital artifacts, even if they may be concerned about the collection and use of personal data

see discussions about the privacy paradox

(6) Many people appear to be resigned to data practices: privacy fatigue, privacy helplessness, privacy apathy, …

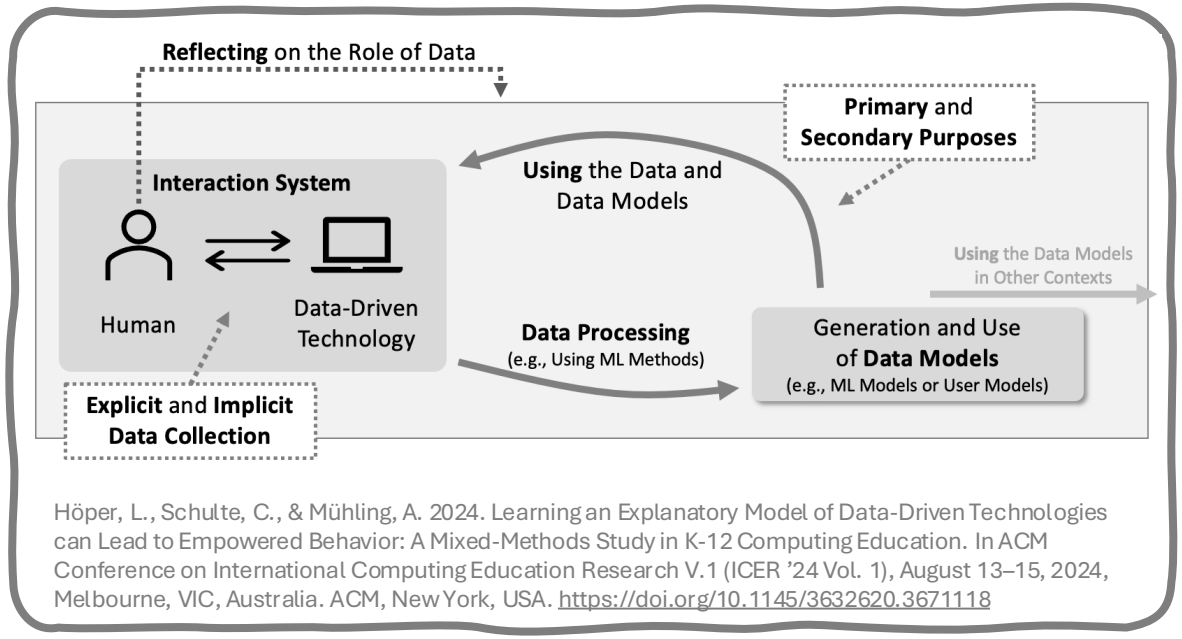
See overview in: Draper, N. A., Hoffmann, C. P., Lutz, C., Ranzini, G., & Turow, J. (2024). Privacy resignation, apathy, and cynicism: Introduction to a special theme. Big Data & Society, 11(3).

Challenges for these goals

1. Knowledge and skills to understand and reflect on the role of data in applications
2. Knowledge and skills learned in class should be meaningful and applicable in such situations

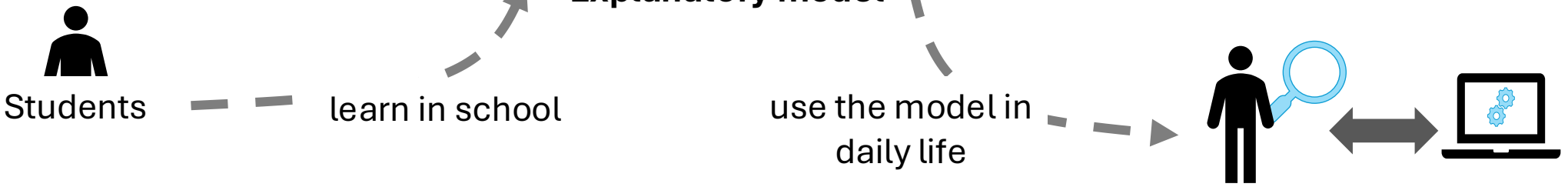
How can we reach these goals?

Idea for developing data awareness



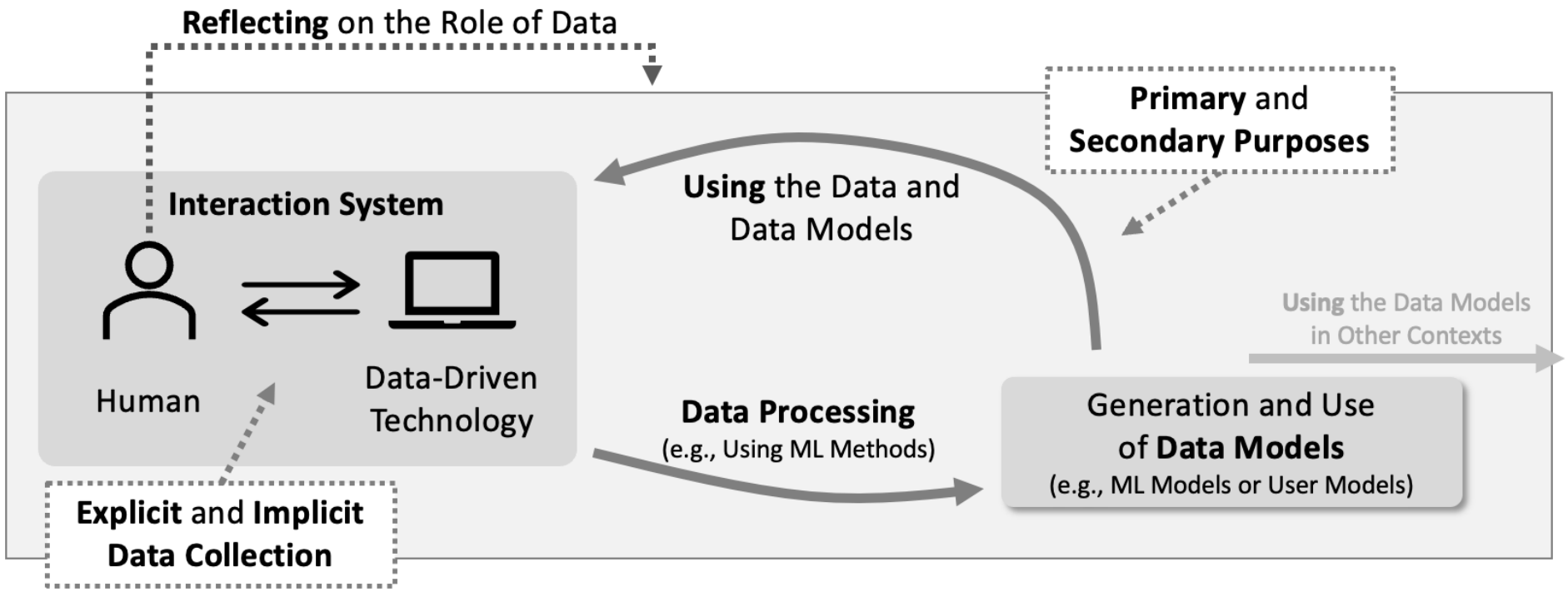
Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118

Explanatory model



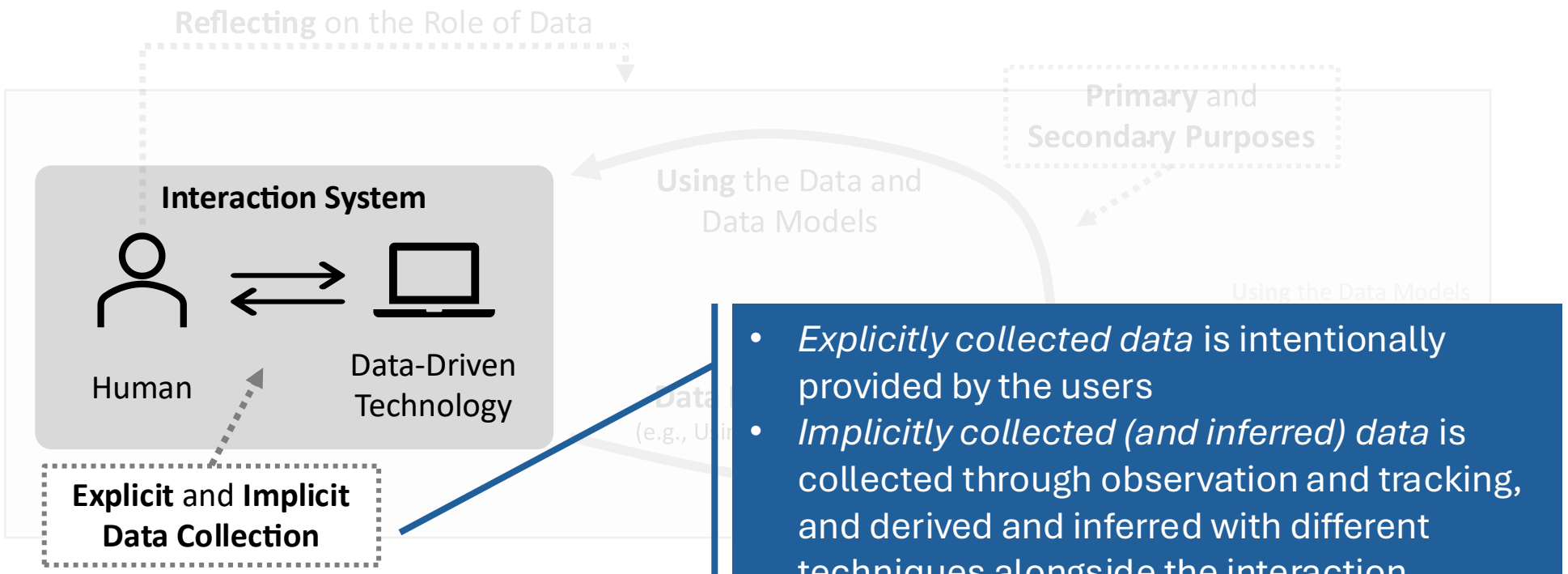
An explanatory model for data practices embedded in data-driven technologies

Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118



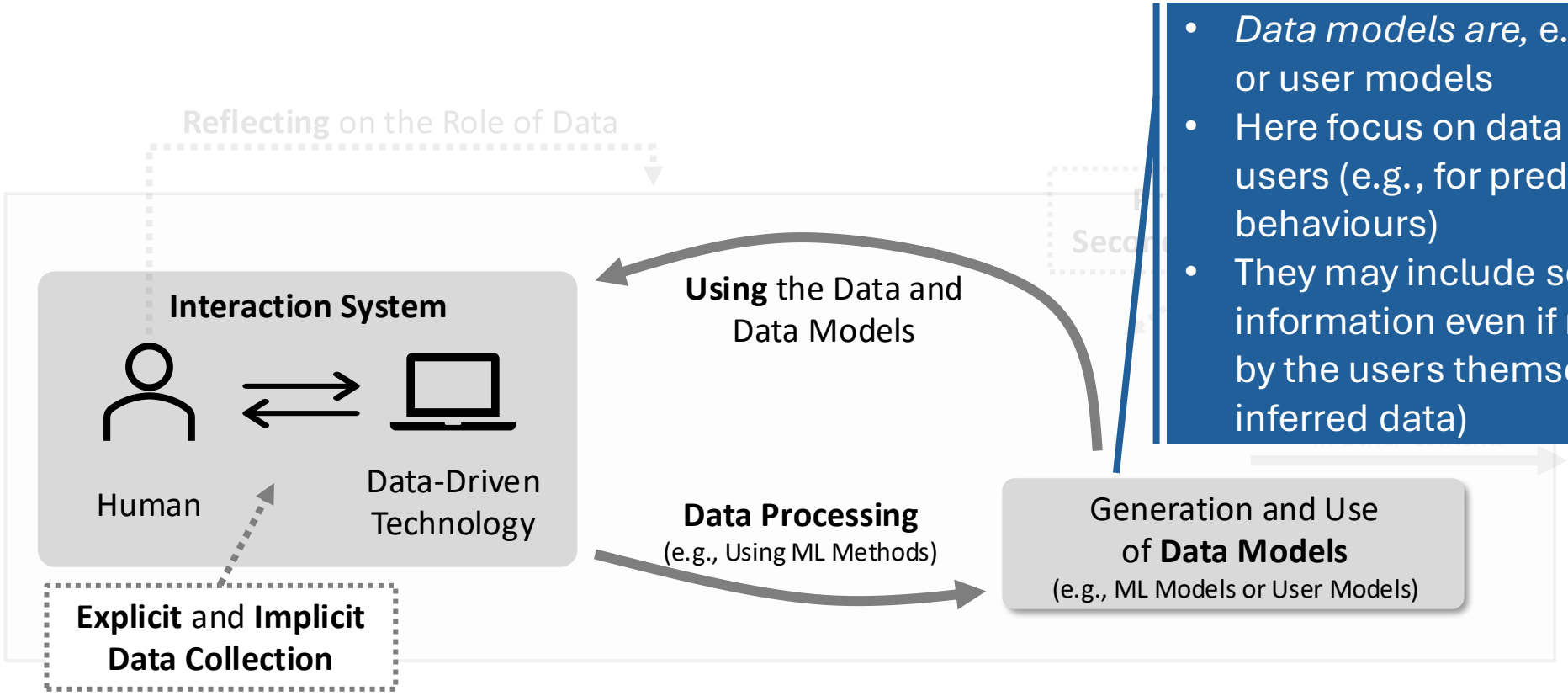
An explanatory model for data practices embedded in data-driven technologies

Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118



An explanatory model for data practices embedded in data-driven technologies

Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118



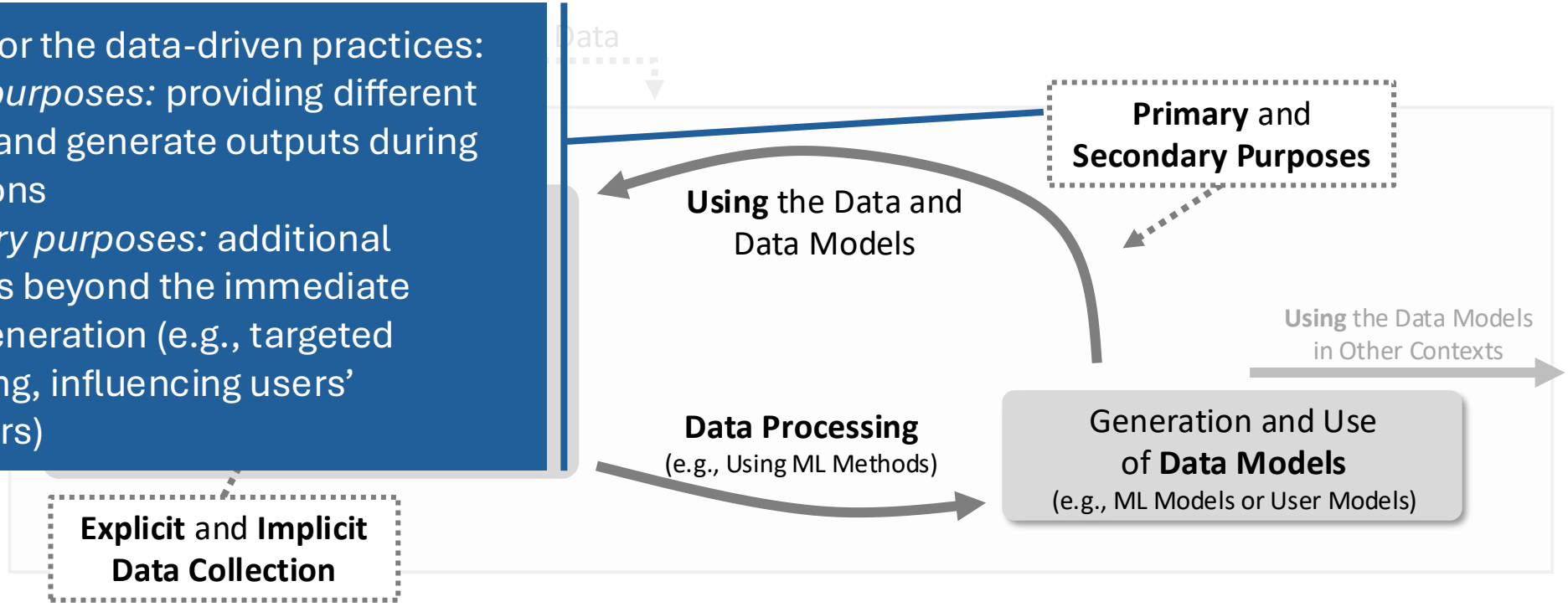
- *Data models are,* e.g., ML models or user models
- Here focus on data models about users (e.g., for predicting behaviours)
- They may include sensible information even if not provided by the users themselves (e.g., inferred data)

An explanatory model for data practices embedded in data-driven technologies

Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118

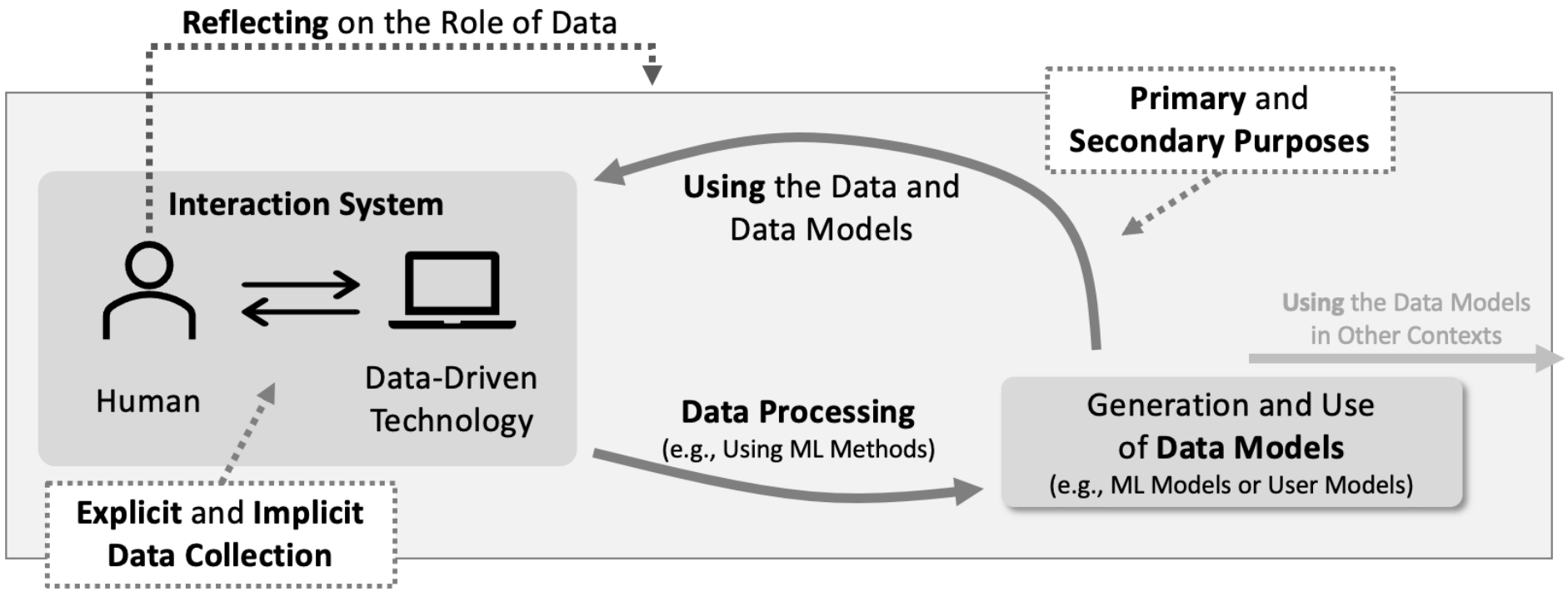
Underlying for the data-driven practices:

- *Primary purposes:* providing different features and generate outputs during interactions
- *Secondary purposes:* additional intentions beyond the immediate output generation (e.g., targeted advertising, influencing users' behaviours)



An explanatory model for data practices embedded in data-driven technologies

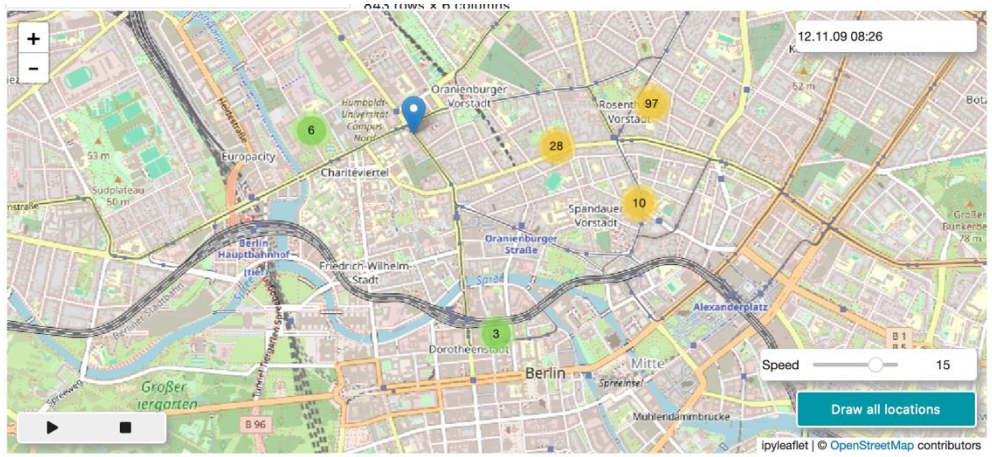
Höper, L., Schulte, C., & Mühling, A. 2024. Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. In ACM Conference on International Computing Education Research V.1 (ICER '24 Vol. 1), August 13–15, 2024, Melbourne, VIC, Australia. ACM, New York, USA. https://doi.org/10.1145/3632620.3671118

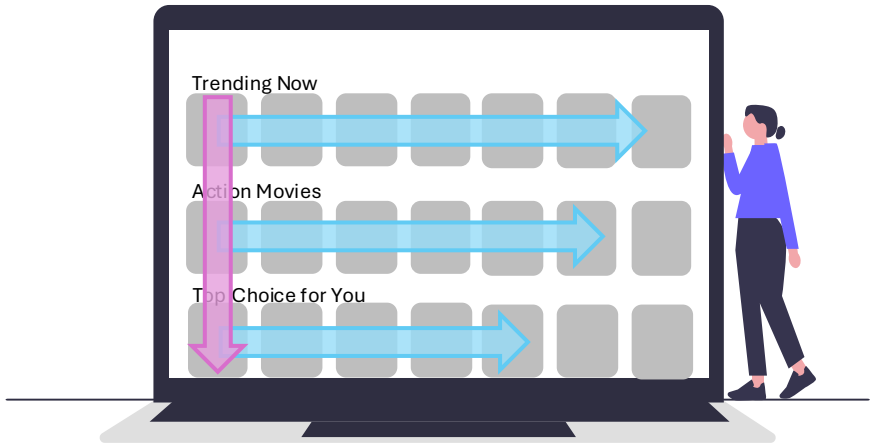


Teaching data awareness in schools

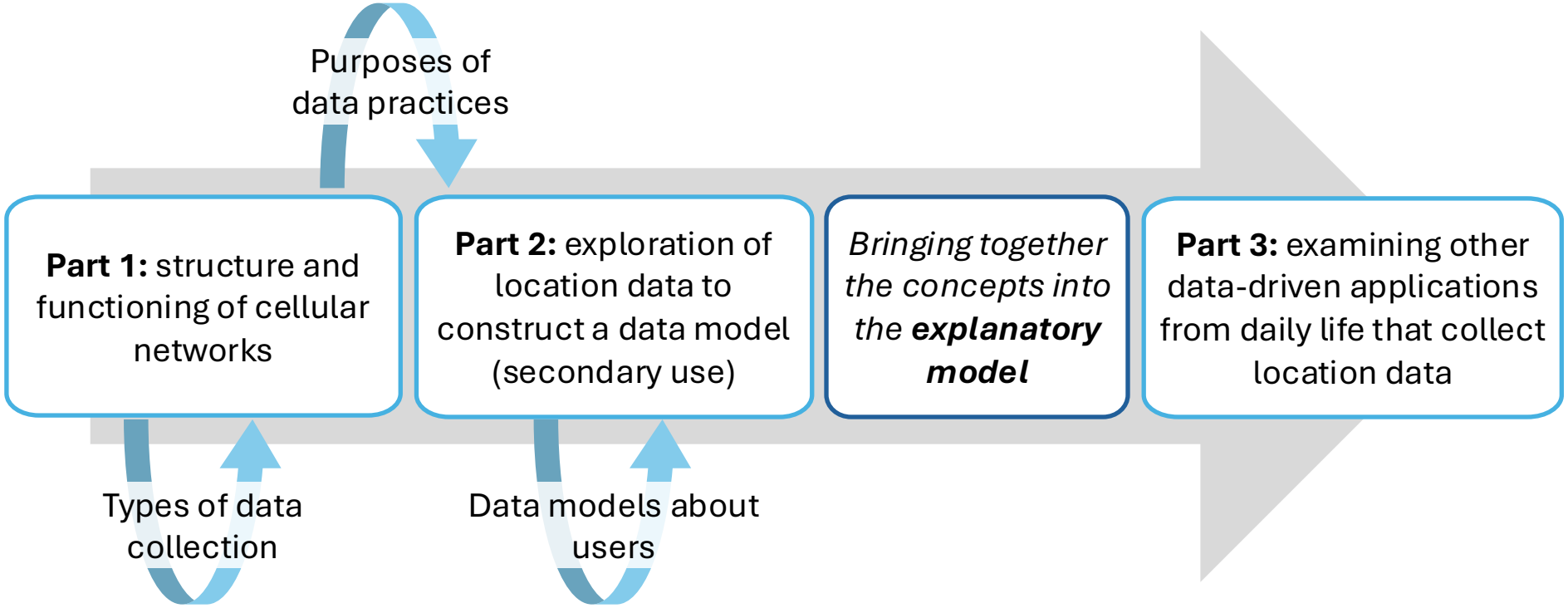
Teaching Unit 1:
Location data and cellular networks
(incl. analysing real location data from a cellular network provider)

Teaching Unit 2:
Recommendation systems in the context of streaming services
(incl. making recommendations with real movie rating data from many people)





Teaching unit: Location data and cellular networks



Designed for 4–6 lessons of 45 minutes each, from grade 6 onwards

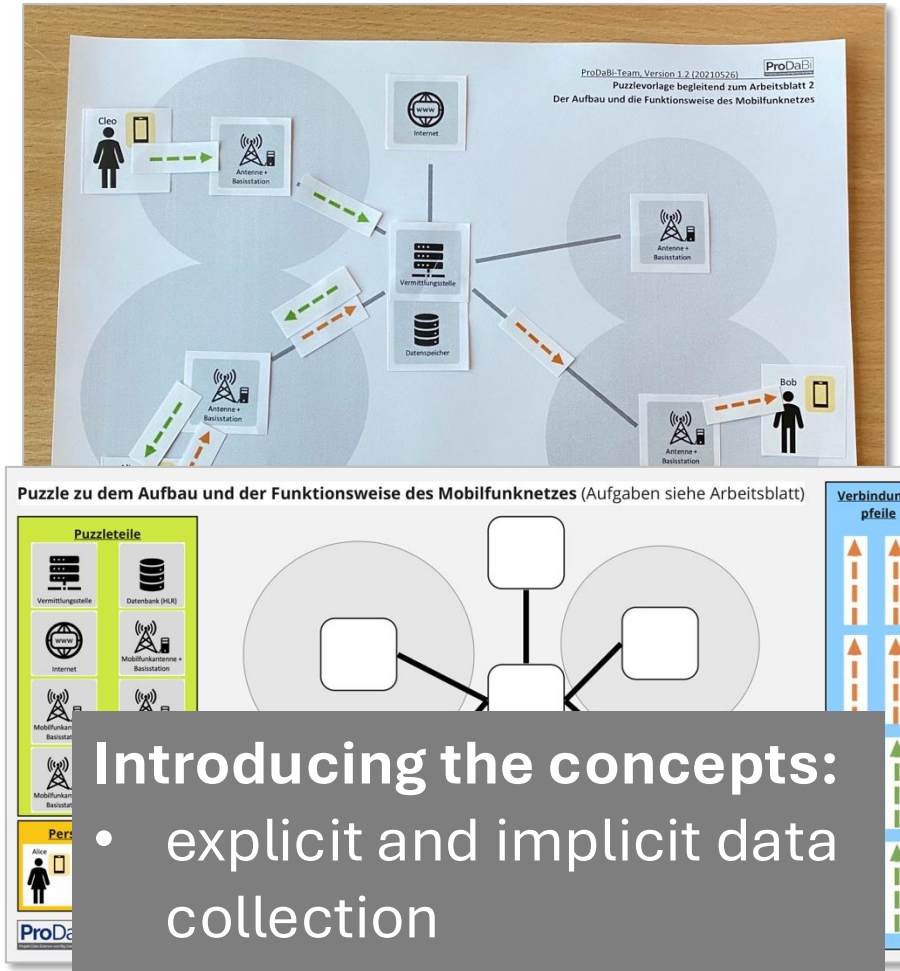
Part 1: structure and functioning of cellular networks

Introduction and activation of personal experiences:

- Thought experiment: establishing a mobile phone connection between two devices
- Students explore the main components of a cellular network using an explanatory video

Puzzle on the structure and functioning of cellular networks:

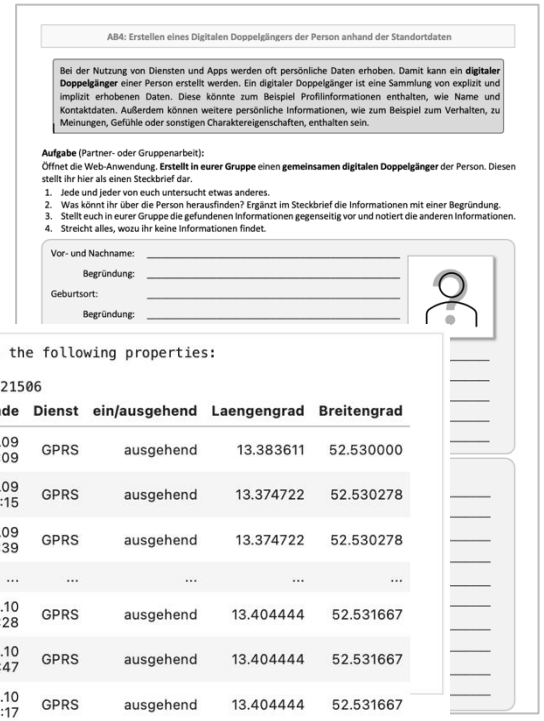
- Simulation of establishing a phone connection
- Identifying key steps and developing ideas about the collected data (→ the relevance of location data becomes clear)



Introducing the concepts:

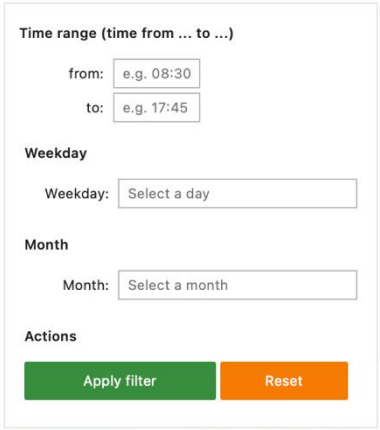
- explicit and implicit data collection
- primary and secondary purposes

Part 2: exploration of location data to construct a data model

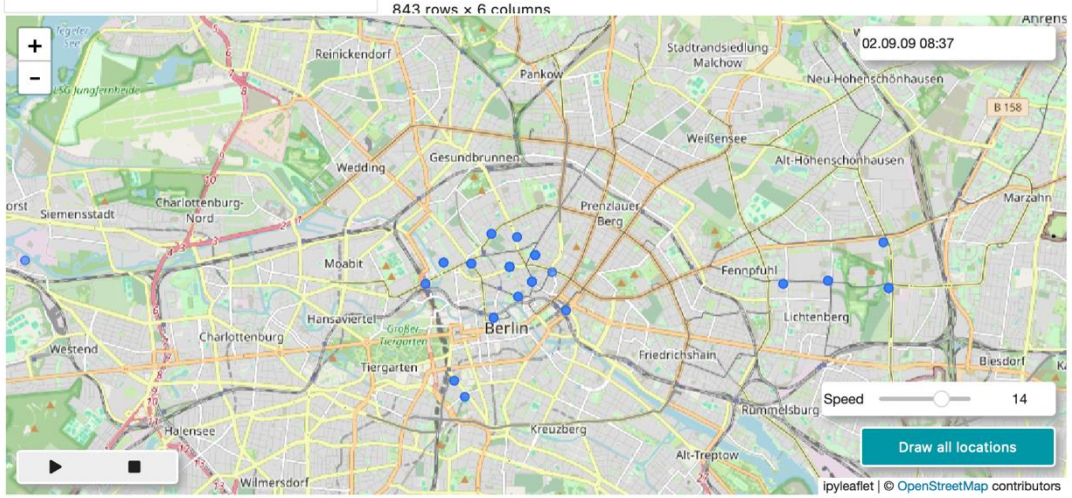


Students explore location data from the cellular network to create a profile (as a fictional secondary purpose):

- **Data:** real cellular network data from one person over half a year (politician Malte Spitz), incl. location data (see: https://www.zeit.de/digital/datenschutz/2011-02/vorratsdaten-malte-spitz)
- **Tool:** a self-developed application embedded in a Jupyter notebook that is accessible via a web browser (filtering data, exploring the data on a map)
- **Task:** Extract as much information as possible about the person behind the data and create a profile



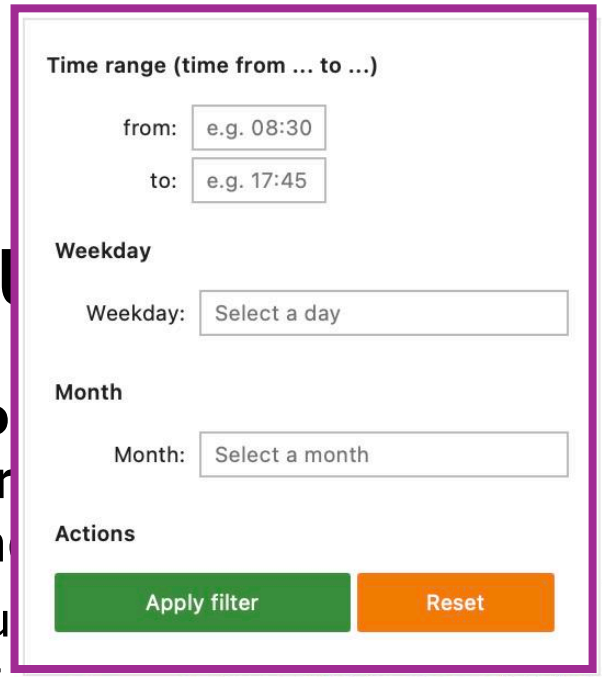
	Beginn	Ende	Dienst	ein/ausgehend	Laengengrad	Breitengrad
0	31.08.09 08:09	31.08.09 08:09	GPRS	ausgehend	13.383611	52.530000
1	31.08.09 08:09	31.08.09 08:15	GPRS	ausgehend	13.374722	52.530278
2	31.08.09 08:15	31.08.09 08:39	GPRS	ausgehend	13.374722	52.530278
...	...	...	...	...	...	...
840	12.02.10 08:58	12.02.10 09:28	GPRS	ausgehend	13.404444	52.531667
841	13.02.10 08:17	13.02.10 08:47	GPRS	ausgehend	13.404444	52.531667
842	13.02.10 08:47	13.02.10 09:17	GPRS	ausgehend	13.404444	52.531667

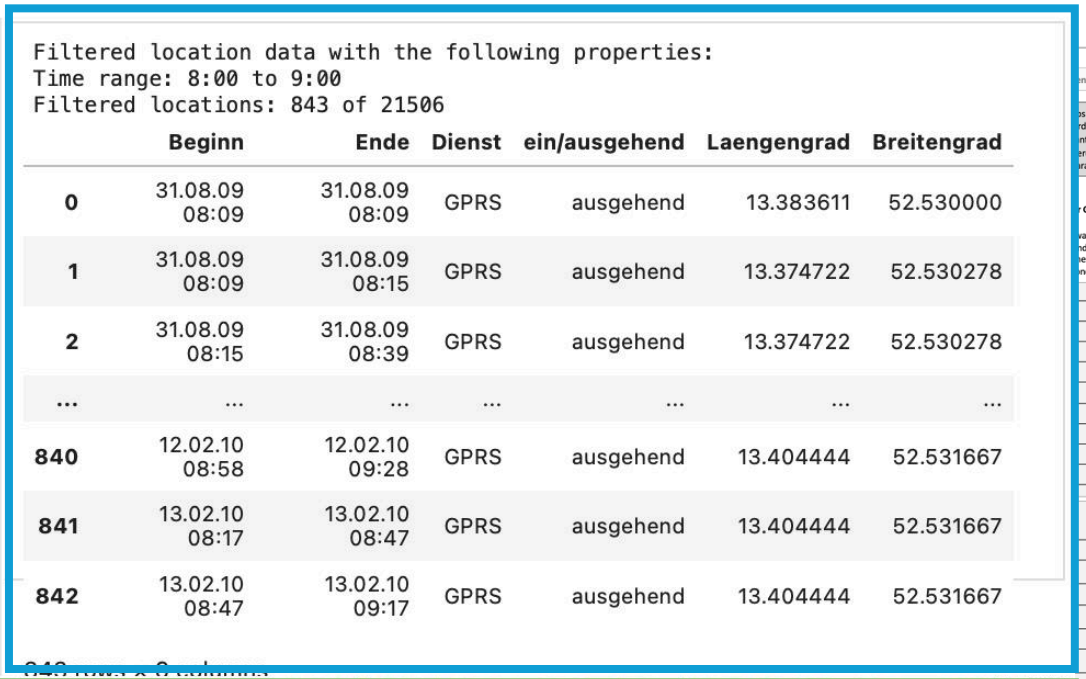


Part 2: constru

Students explo cellular networ fictional secon

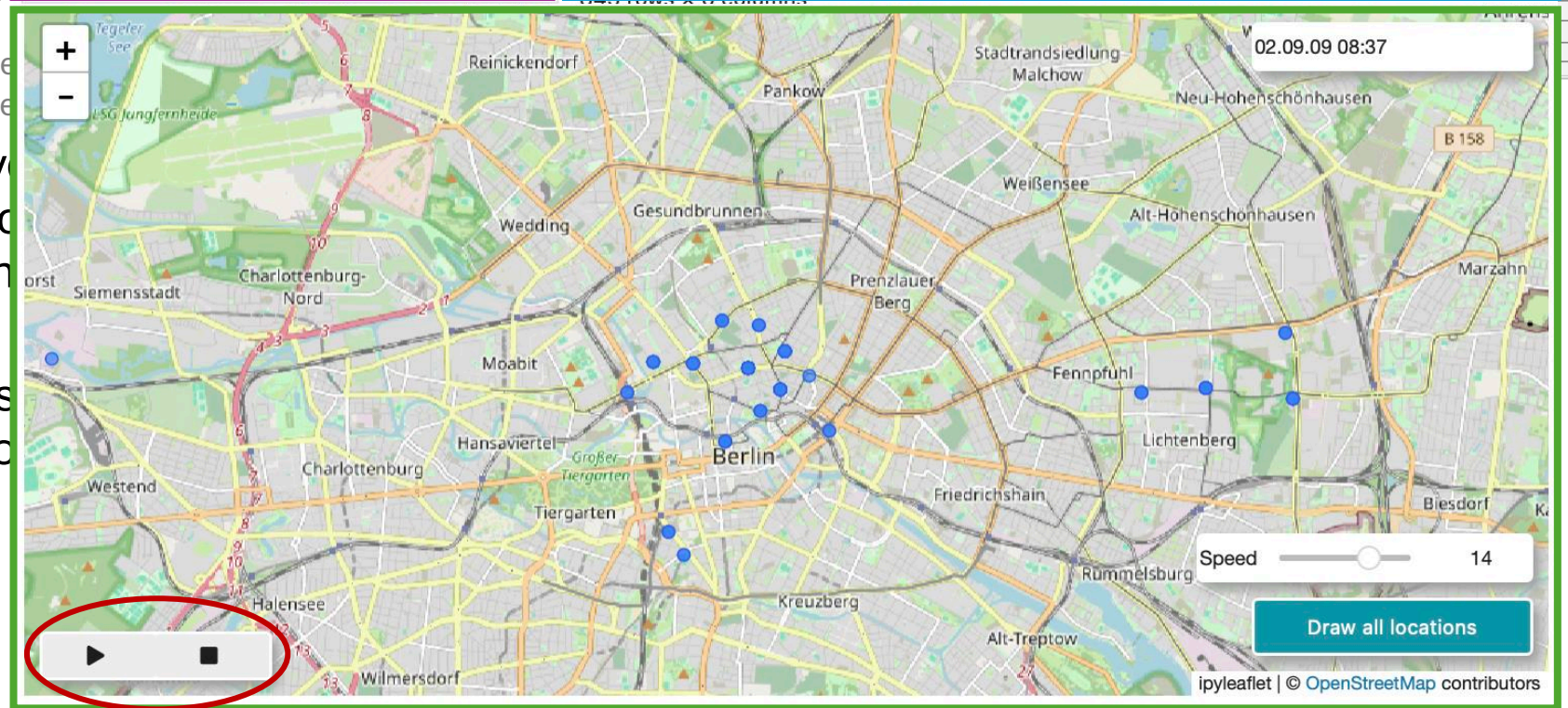
- **Data:** real cellu over half a year location data (se 02/vorratsdaten-malte
- **Tool:** a self-dev Jupyter noteboo browser (filterin map)
- **Task:** Extract as about the perso profile





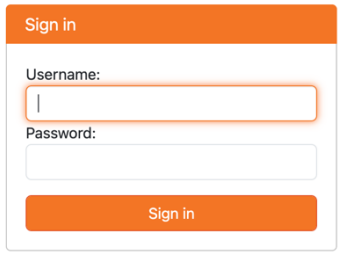
Filtered location data with the following properties:
Time range: 8:00 to 9:00
Filtered locations: 843 of 21506

	Beginn	Ende	Dienst	ein/ausgehend	Laengengrad	Breitengrad
0	31.08.09 08:09	31.08.09 08:09	GPRS	ausgehend	13.383611	52.530000
1	31.08.09 08:09	31.08.09 08:15	GPRS	ausgehend	13.374722	52.530278
2	31.08.09 08:15	31.08.09 08:39	GPRS	ausgehend	13.374722	52.530278
...	...	...	...	...	...	...
840	12.02.10 08:58	12.02.10 09:28	GPRS	ausgehend	13.404444	52.531667
841	13.02.10 08:17	13.02.10 08:47	GPRS	ausgehend	13.404444	52.531667
842	13.02.10 08:47	13.02.10 09:17	GPRS	ausgehend	13.404444	52.531667

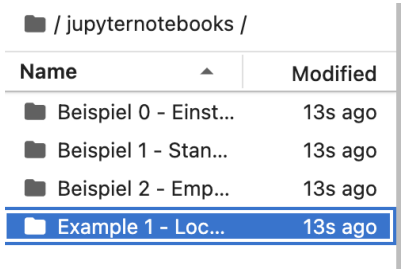


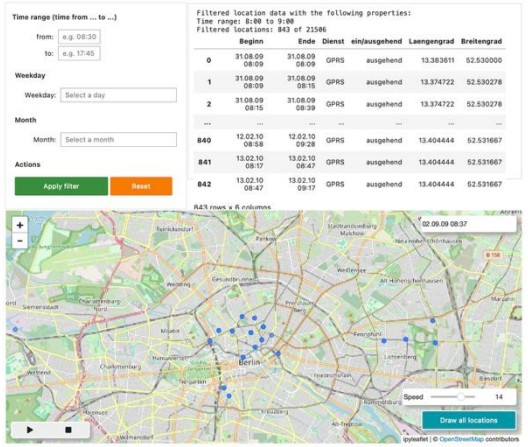
„Activity break“

1. Open the link: https://go.upb.de/ws-db
2. Sign in (choose a name and a password)

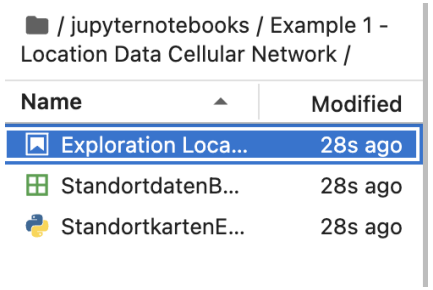


3. Open the folder Example 1

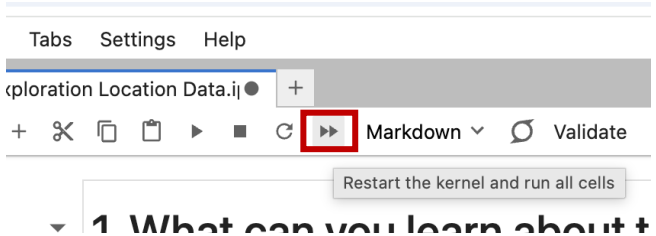




4. Start the Jupyter Notebook „Exploration Location Data“



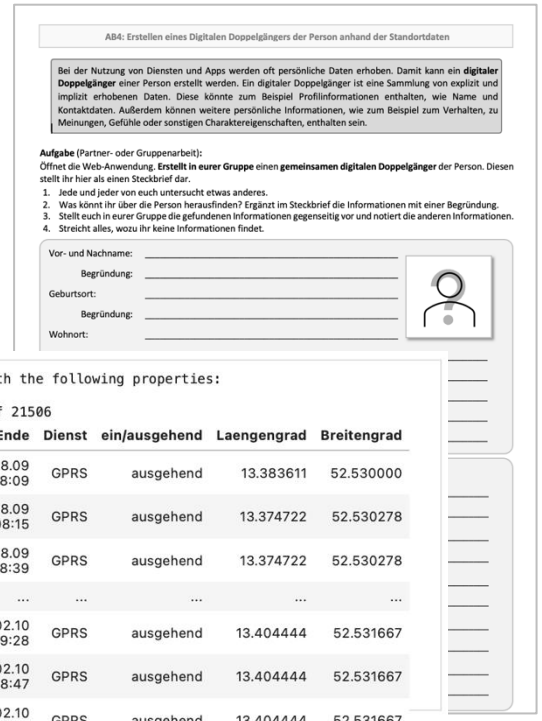
5. Run all cells



6. Load the data (Button in the middle of the screen

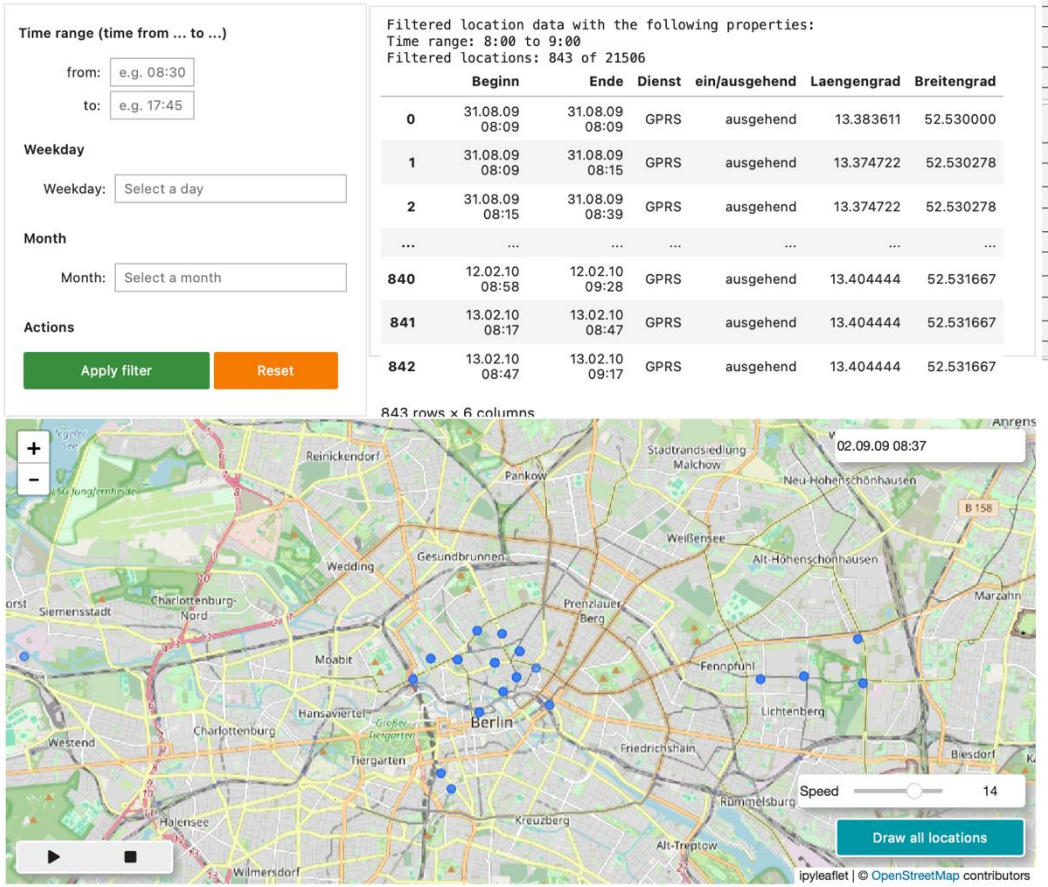
Task:
Explore the data and find out as much as possible about the person. Note down your observations.

Part 2: exploration of location data to construct a data model



Discussion and reflection:

- Students' findings are collected and discussed, including different interpretations
- evaluating these data practices, including the legal regulations surrounding such data (there are good reasons to collect these data, but also regulations how they can be used)



Introducing the concept:

- Data models about users

Results from my students

Lives in Berlin

Family is likely located near Münster

Appears to be involved in politics (frequently present in the government district)

Frequently travels by train within Germany

Occasionally flies within Germany

Often visits restaurants and bars („has money“)

Likely has children (frequent visits to a kindergarten)

Assumptions about occupation: politician, train driver, pilot, insurance employee

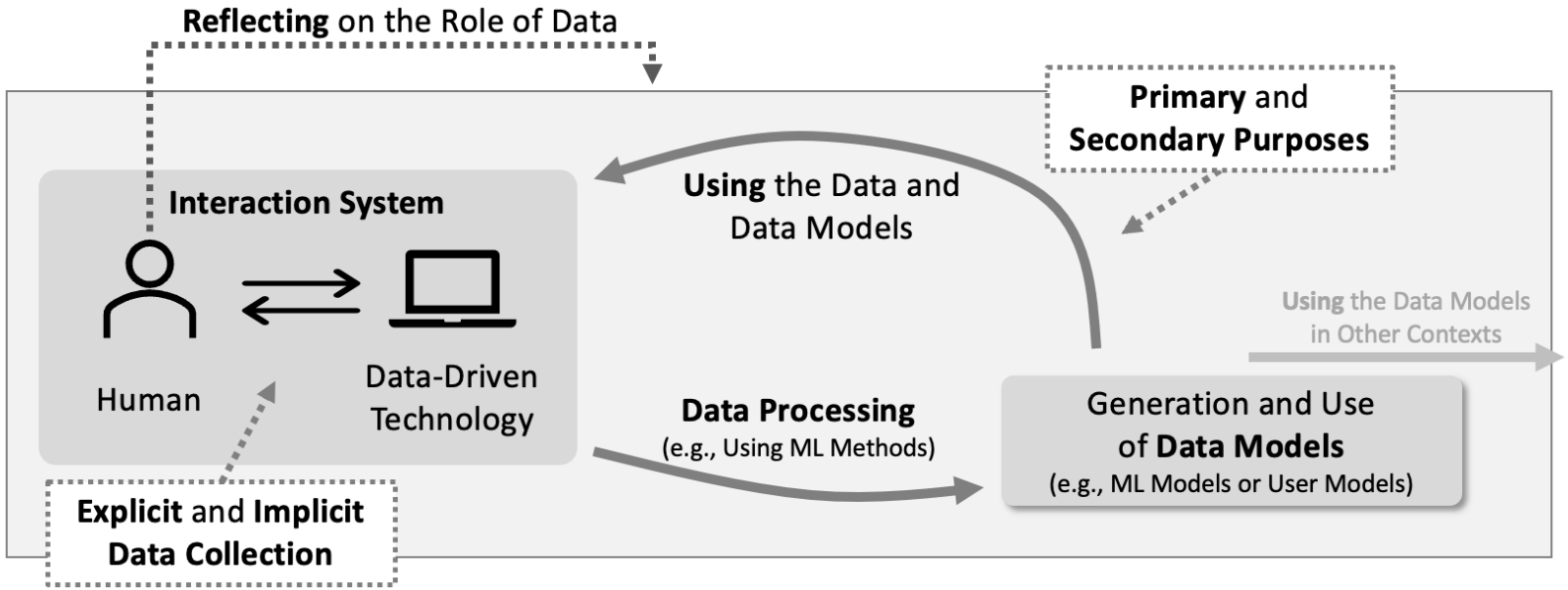
No fixed working hours

Often assumed to be male(?)

Bringing together the concepts into the explanatory model

Developing the explanatory model:

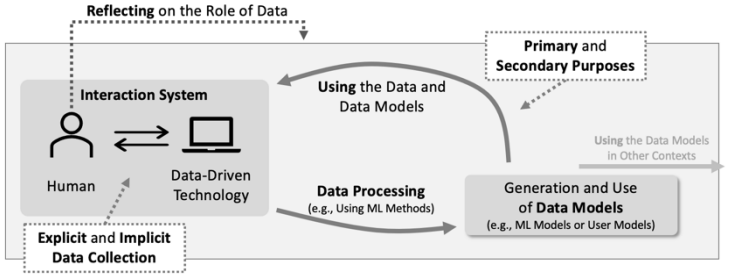
- The concepts learned are combined into the explanatory model
- Students see this model:

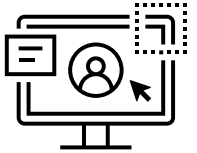


Part 3: applying the explanatory model

Exploring additional everyday contexts:

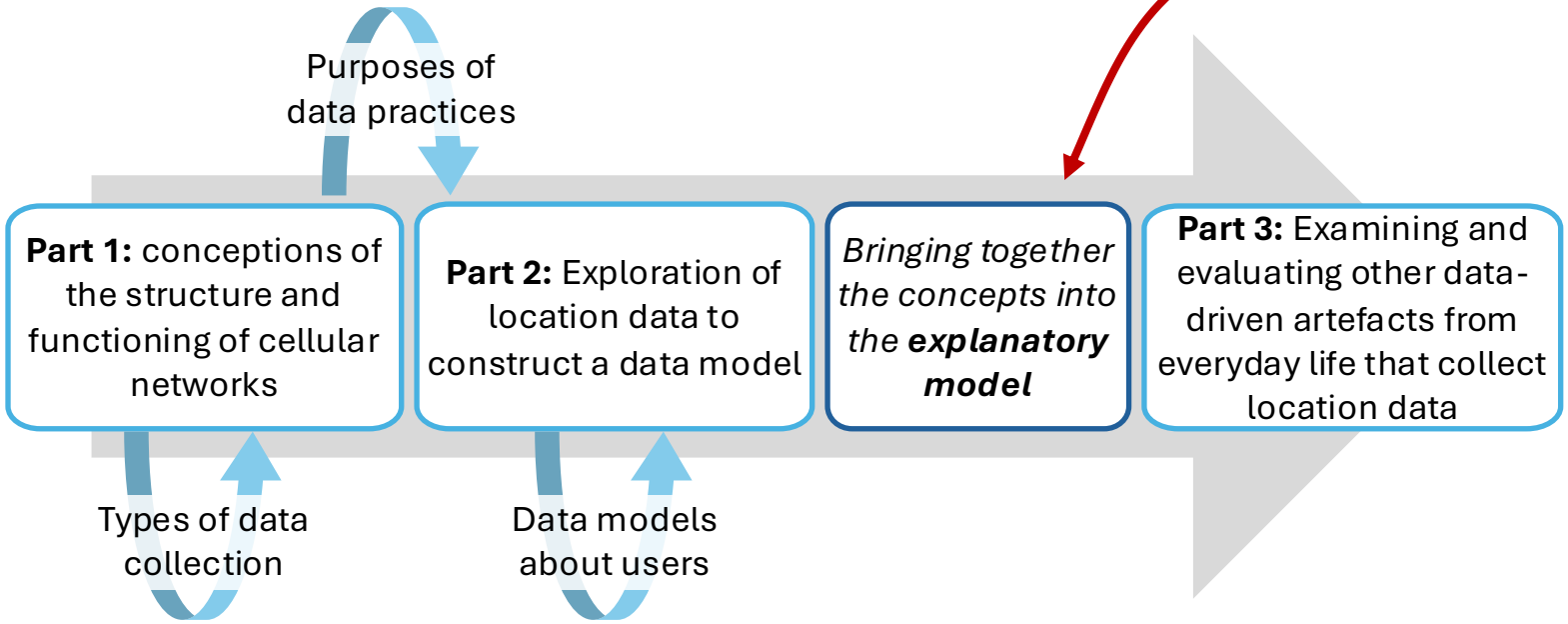
- Students identify further everyday contexts with location data
- They examine different examples using the **explanatory model as an analytical lens**
- They discuss their experiences
- They evaluate the data practices in the different examples and reflect on their own role in these contexts

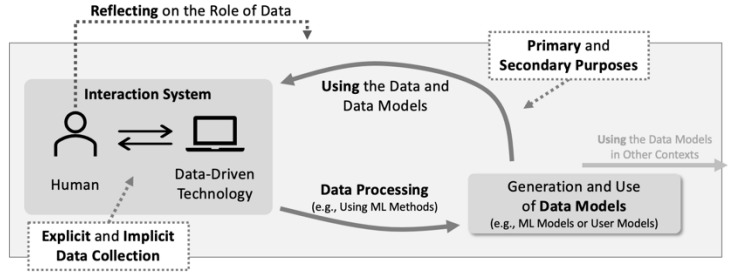




Data-driven applications from daily life

Interesting observation during an empirical study





explanatory model was **designed for teachers** to understand data awareness

Based on experiences and research findings, the explanatory model proved to be the **learning topic for students**

Results from a design-based research project

Over a design-based research project we have found:

- Students (from 12 years) understand the aspects of the explanatory model and can use it uncover and evaluate data-driven practices in everyday applications
- This can support students in taking a broader view on the role of data in these systems

(see: Höper, L., & Schulte, C. (2024). The data awareness framework as part of data literacies in K-12 education. Information and Learning Sciences, 125(7/8), 491–512. [Link])

- Learning the explanatory model encourages them to relate it to their everyday experiences and reflect on their daily interactions with these technologies

(see: Höper, L., & Schulte, C. (2024). Empowering Students for the Data-Driven World: A Qualitative Study of the Relevance of Learning about Data-Driven Technologies. Informatics in Education, 23(3), 593–624. [Link])

- Learning the explanatory model can support students to become more informed about the role of data in data-driven technologies and lead to more empowered behaviour

(see: Höper, L., Schulte, C., & Mühling, A. (2024). Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. Proceedings of the 2024 ACM Conference on International Computing Education Research – V. 1 (ICER 2024), 326–342. [Link])

Outlook: Explanatory model as a tool for LLM-based analysis as a starting point?

I have the following analytical lens from an explanatory model for the role of data in data-driven applications:
Explicitly collected data is intentionally provided by the users
Implicitly collected (and inferred) data is collected through observation and tracking, and derived and inferred with different techniques alongside the interaction
Data models about the user (e.g., for predicting behaviours): Based on explicitly and implicitly collected data, a data model about the user is generated, which may include sensible information even if not provided by the user self.
Primary purposes: providing different features and generate outputs during interactions
Secondary purposes: additional intentions beyond the immediate output generation. This also includes problematic secondary use of data.

You are now an analyst who uses this framework to analyze daily digital applications. I give you an application. For each of the five analysis aspects, write down your results in bullet points. Keep the bullet points as short as possible; do not write long texts or explanations. Provide sources where you have your information from.

Weniger anzeigen

Prompt / GPT-5.3 mini

1) Explicitly collected data

- Search queries typed or spoken
- Voice input (if voice search used)
- Language preferences (selected or inferred)
- Location input (e.g., region, city settings)
- Feedback signals (e.g., "Was this helpful")
- Account data (if signed in: Google account info)

Sources:
Google Privacy Policy ↗
Google Search Help / Data & Privacy ↗

2) Implicitly collected / inferred data

- Clicked results and click-through behavior
- Dwell time on pages (how long results are viewed)
- Scroll depth and interaction patterns
- Query reformulation sequences
- Device information (OS, browser, hardware)
- Approximate location via IP / GPS (if enabled)
- Browsing activity via Google services (if Web & App Activity is enabled)
- Inferred interests from aggregated behavior

Sources:
Google Privacy Policy (Data Collection & Usage) ↗
Google Technologies & Partner Sites ↗

3) Data models about the user

- Interest profiles (topics, hobbies, intent categories)
- Search personalization models (ranking adjustments per user)
- Location-based preference modeling (local intent, mobility patterns)
- Ad personalization profiles linked via Google ecosystem
- Behavioral prediction signals (next query / likely intent)
- Potential sensitive inferences (e.g., health, finance, politics inferred from queries)

Sources:
Google Ads Personalization Overview ↗
Google Privacy Policy ↗

4) Primary purposes

- Web search and information retrieval
- Ranking and organizing search results
- Autocomplete and query suggestions
- Providing direct answers (snippets, knowledge panels)
- Local search results (maps, nearby services)
- Improving relevance and speed of results

Sources:
How Google Search Works ↗
Google Search Help ↗

5) Secondary purposes (beyond immediate search output)

- Targeted advertising across Google services
- Cross-service personalization (YouTube, Maps, Gmail ecosystem)
- Product improvement and model training (search quality, AI systems)
- User profiling for analytics and engagement optimization
- Fraud/spam detection and platform integrity
- Commercial ecosystem optimization (ad revenue prioritization)

Sources:
Google Ads & Data Usage ↗
Google Privacy Policy ↗

This could also be done within the teaching unit and serve as a starting point for discussing the mentioned practices.

Further literature on data awareness

- Höper, L. (2025). *Entwicklung und Evaluation des Konzepts Datenbewusstsein für informatische Bildung* [PhD Thesis, Paderborn University]. https://digital.ub.uni-paderborn.de/doi/10.17619/UNIPB/1-2420
- Höper, L., & Schulte, C. (2025). ReVEAL model and its application to revealing viewpoints on educational approaches to learning about data and AI. *Computer Science Education*, 1–33. https://doi.org/10.1080/08993408.2025.2516957
- Höper, L., Schulte, C., & Mühling, A. (2024). Learning an Explanatory Model of Data-Driven Technologies can Lead to Empowered Behavior: A Mixed-Methods Study in K-12 Computing Education. *Proceedings of the 2024 ACM Conference on International Computing Education Research - Volume 1 (ICER 2024)*, 326–342. https://doi.org/10.1145/3632620.3671118
- Höper, L., & Schulte, C. (2024). Empowering Students for the Data-Driven World: A Qualitative Study of the Relevance of Learning about Data-Driven Technologies. *Informatics in Education*, *23*(3), 593–624. https://doi.org/10.15388/infedu.2024.19
- Höper, L., & Schulte, C. (2024). The data awareness framework as part of data literacies in K-12 education. *Information and Learning Sciences*, *125*(7/8), 491–512. https://doi.org/10.1108/ILS-06-2023-0075