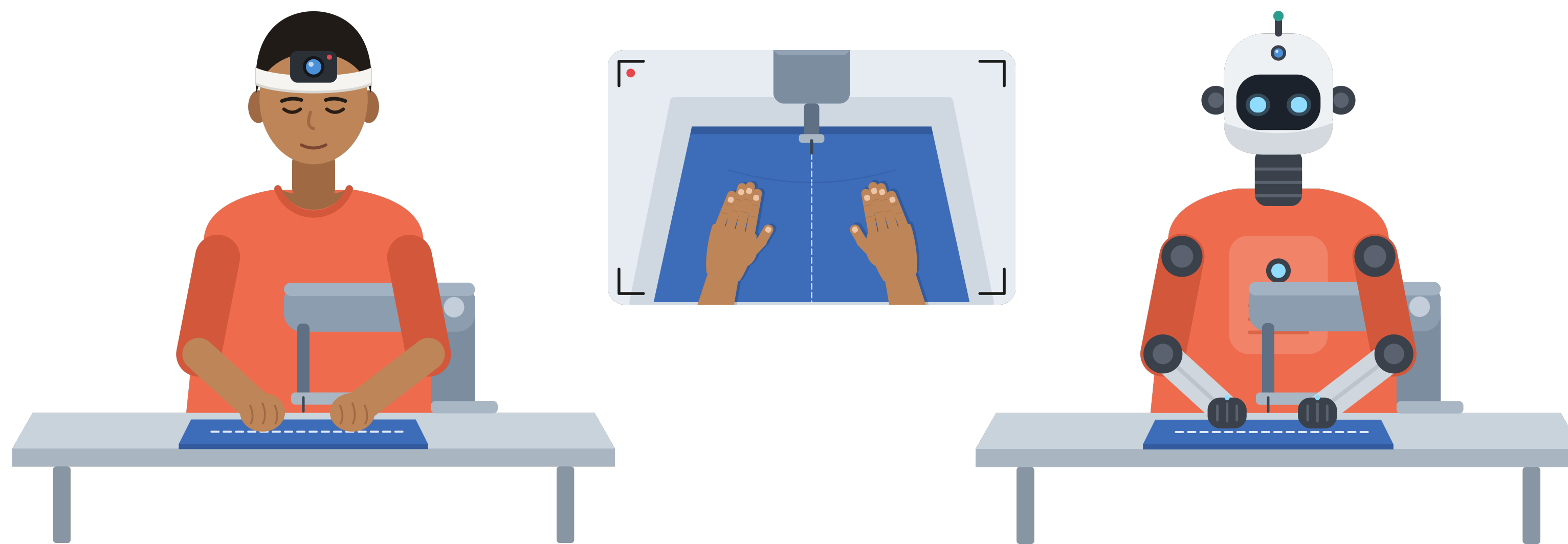


Recording for Robots

The Labour Behind Egocentric Data Collection

Sophia N. Wilson, Department of Computer Science, University of Copenhagen, Denmark

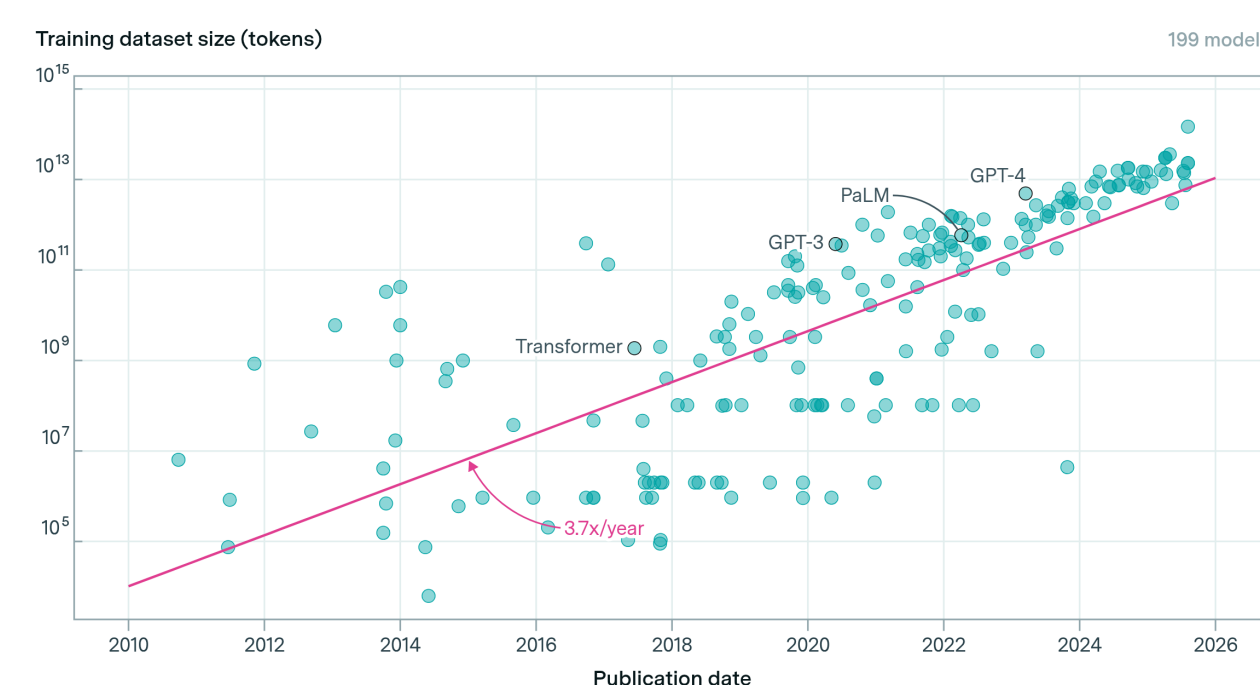
Robots increasingly learn from first-person video recorded with head-mounted cameras. Producing this data is a new form of labour. Through interviews and surveys with data collectors in India and annotators in Kenya, we examine how workers experience this work and understand its downstream use.



Workers wear head-mounted cameras while doing their usual tasks including work and household chores (left). The first-person footage (centre) becomes egocentric data used to train robots to perform the same work (right). In this way, everyday labour is turned into training data for the systems meant to automate it.

We are shifting from a state where we were building models from data to actively creating data for building models [1]

Hyper-datafication refers to the industrialised production and accumulation of data for AI model development through three processes.



(1) The scaling and recombination of existing data sources [2]



(2) The use of AI systems to generate synthetic data [3]

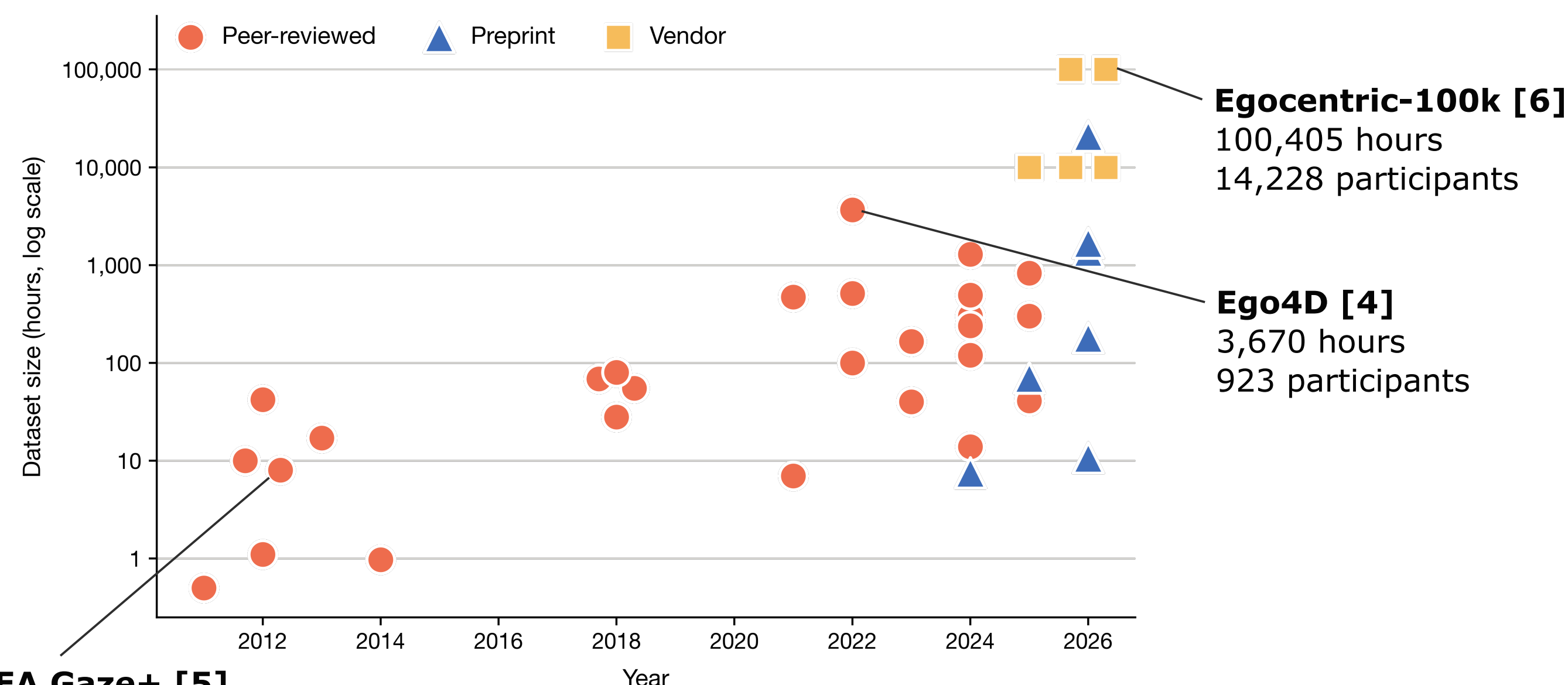


(3) The creation of purpose-built AI data [4]



Hyper-datafication (FACCT 2026)

The size of egocentric datasets has increased by five orders of magnitude since 2012



GTEA Gaze+ [5]
8 hours
10 participants

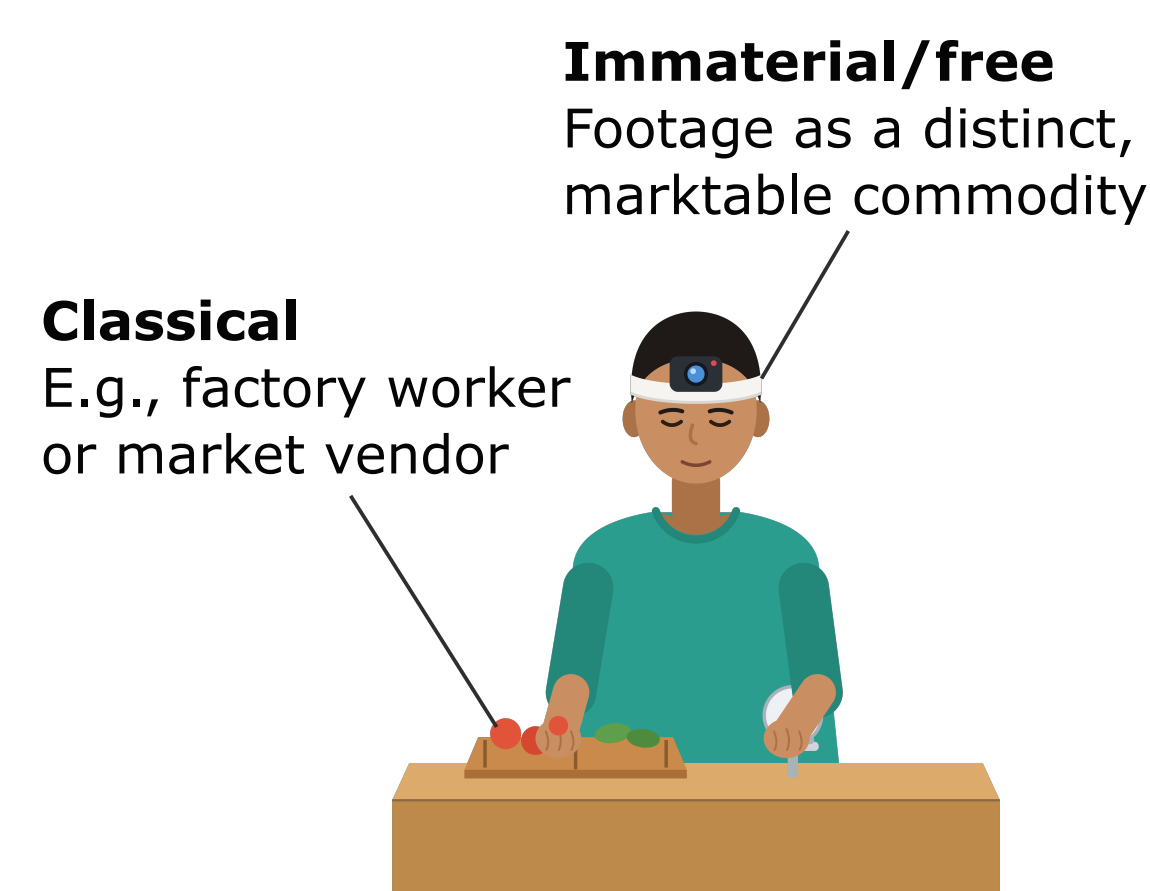
Egocentric data collection has given rise to a new form of labour

We analyse egocentric data collection through the lens of labour, distinguishing:

Classical labour: work producing material value.

Immaterial labour: work that produces informational and cultural value through affective and cognitive activity [7].

Free labour: activities performed for free that other commodify [8].



Research questions

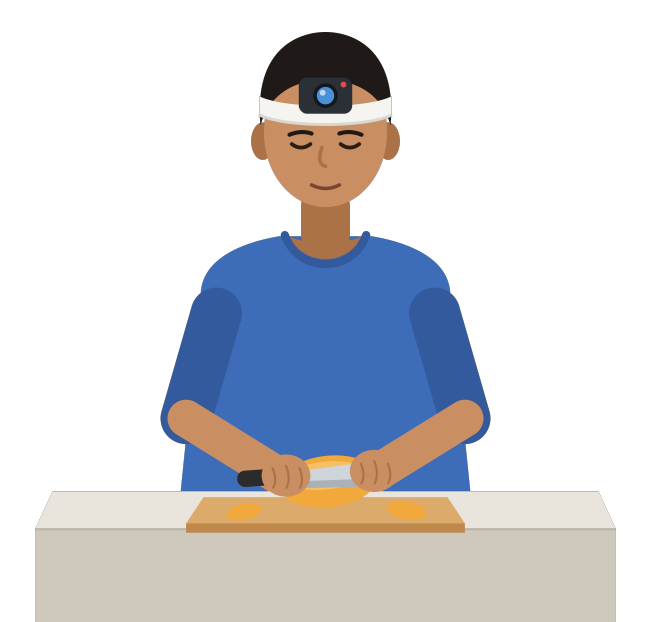
Research objective: This study aims to understand how the workers who collect and annotate egocentric data experience this work, and how they understand the downstream use of the data they produce.

RQ1 How do data collectors perceive the work of egocentric recording?

Q11 Do they see recordings as additional work, or as being paid for tasks they would do anyway?

Q12 How, if at all, does recording change the way they carry out the recorded tasks?

Q13 How are they compensated for recording, and how do they view this compensation?



Data collectors
Bengaluru, India

RQ2 How do data collectors understand the data they produce and its downstream use?

Q21 What do they understand the recordings capture?

Q22 How do they understand the purpose of the data?

Q23 What information are they given about downstream use?



Data collectors
Bengaluru, India

RQ3 How do annotators process egocentric data, and how do they understand its downstream use?

Q31 What content do they encounter in the recordings?

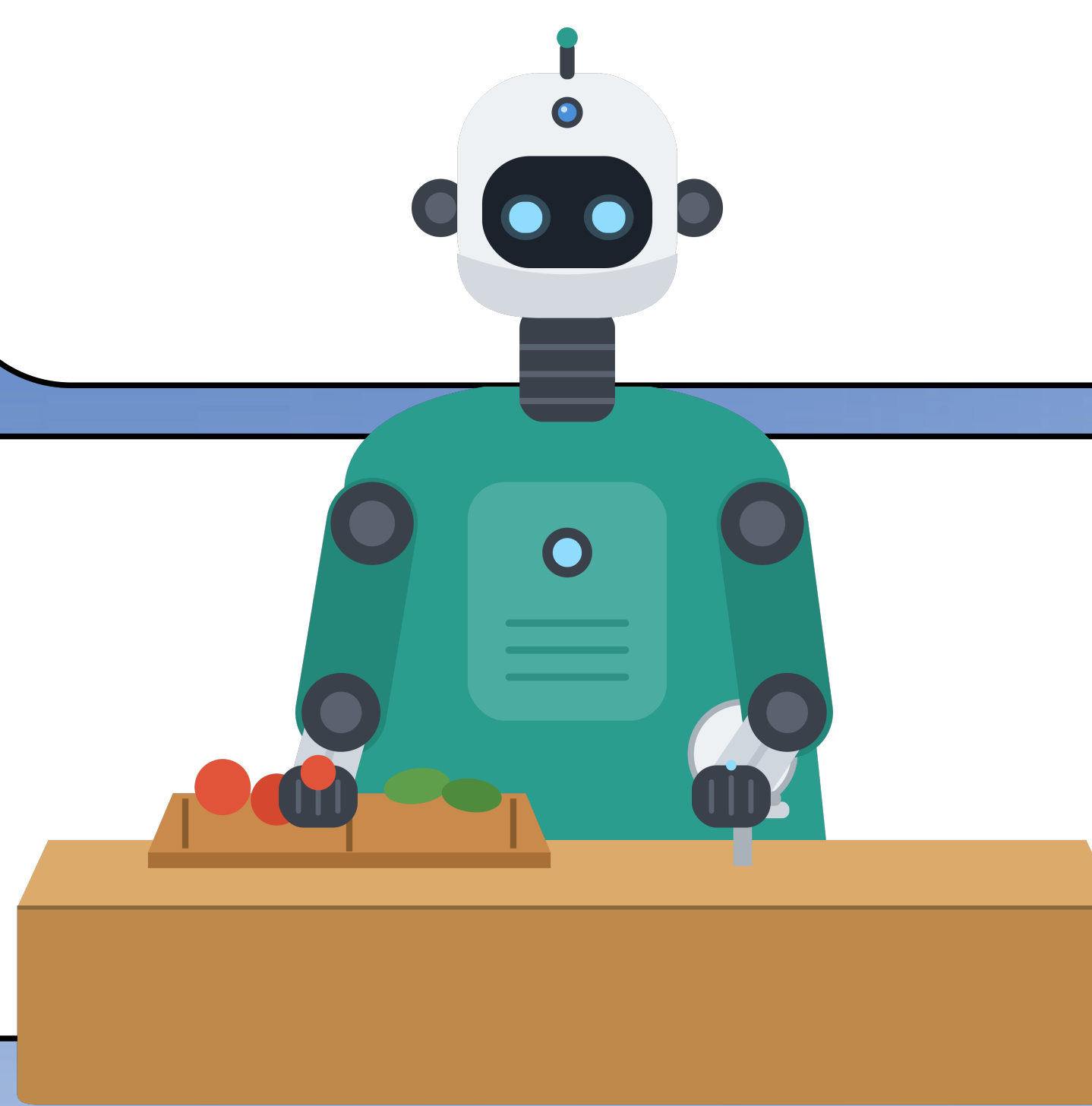
Q32 What are they asked to annotate?

Q33 What do they understand the purpose of the data to be?



Data annotators
Nairobi, Kenya

- [1] Wilson et al. How Hyper-datafication Impacts the Sustainability Costs in Frontier AI. FACCT 2026.
[2] Epoch AI. The size of datasets used to train language models doubles approximately every six months. 2024.
[3] Image from PromptSora. Tokyo Walk. 2024.
[4] Grauman et al. Ego4D: Around the World in 3,000 Hours of Egocentric Video. CVPR 2022.
[5] Fathi et al. Learning to Recognise Daily Actions using Gaze. ECCV 2012.
[6] Build AI. Egocentric-100k. Hugging Face Datasets. 2025.
[7] Lazzarato et al. Immaterial labor. Radical thought in Italy: A potential politics. 1996.
[8] Terranova. Free labor: Producing culture for the digital economy. Social text. 2000.



Examples of
egocentric
data collection