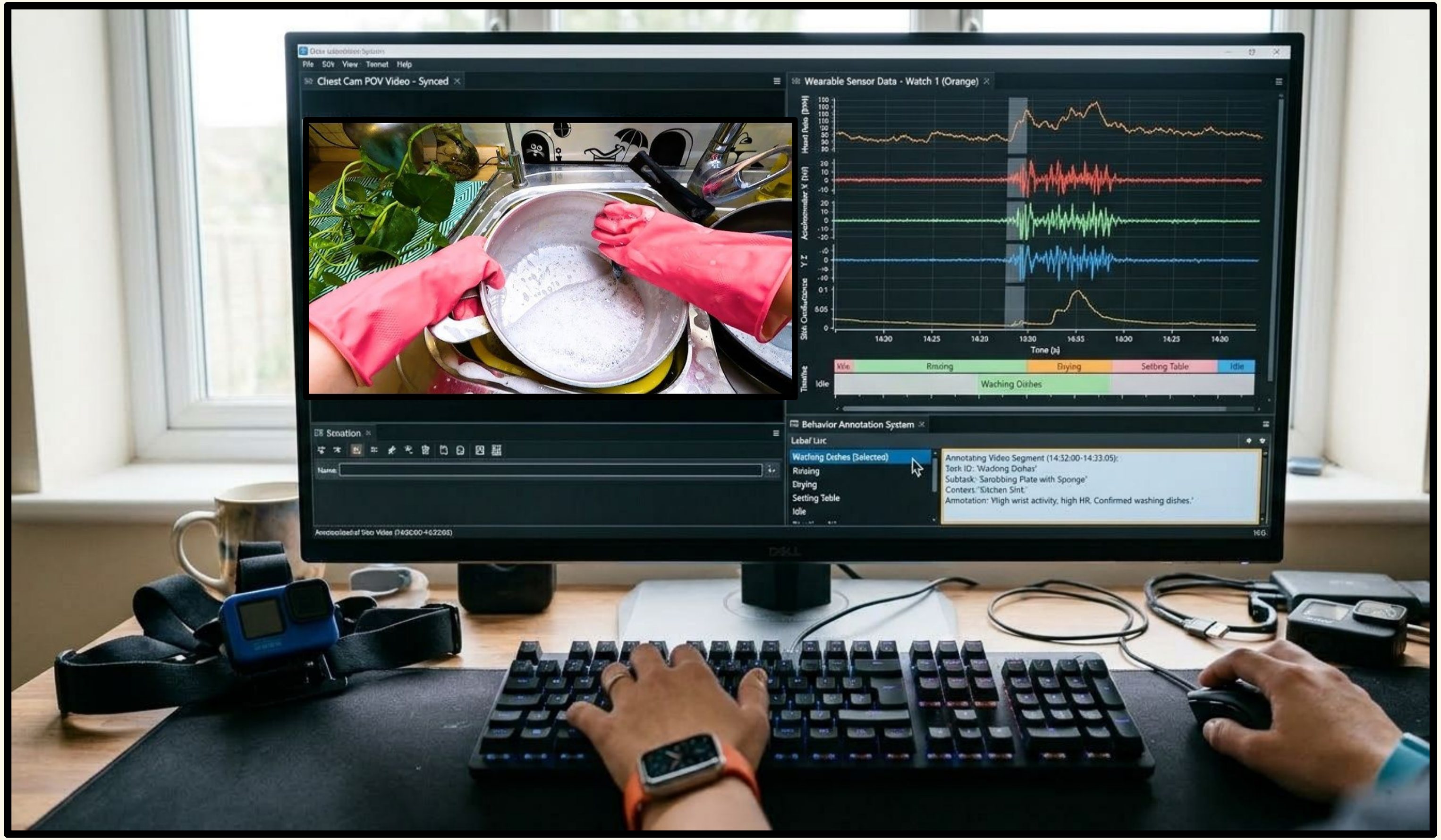
CALIBRATION SLIDE

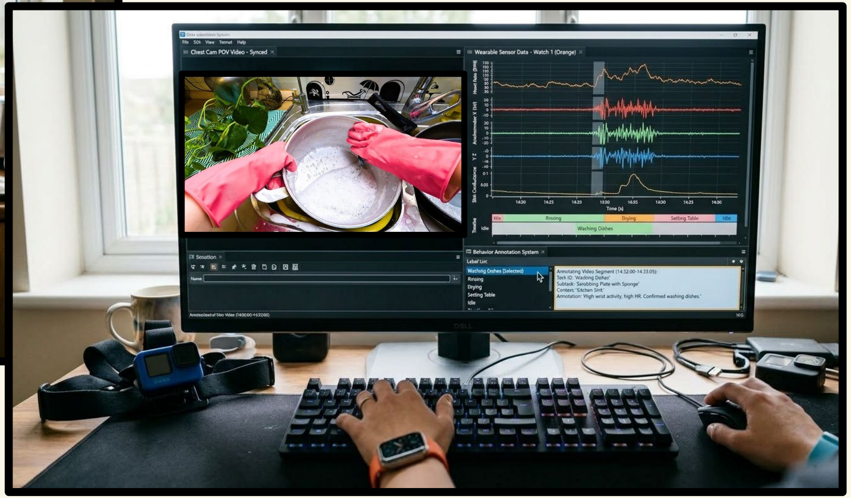
Wrist-based wearable devices

First person view camera

Wrist-based wearable devices



File
View
Help
Chest Cam POV Video - Synced
Wearable Sensor Data - Watch 1 (Orange)
Setting Table
Idle
Behavior Annotation System
Rinsing
Drying
Setting Teble
Idle



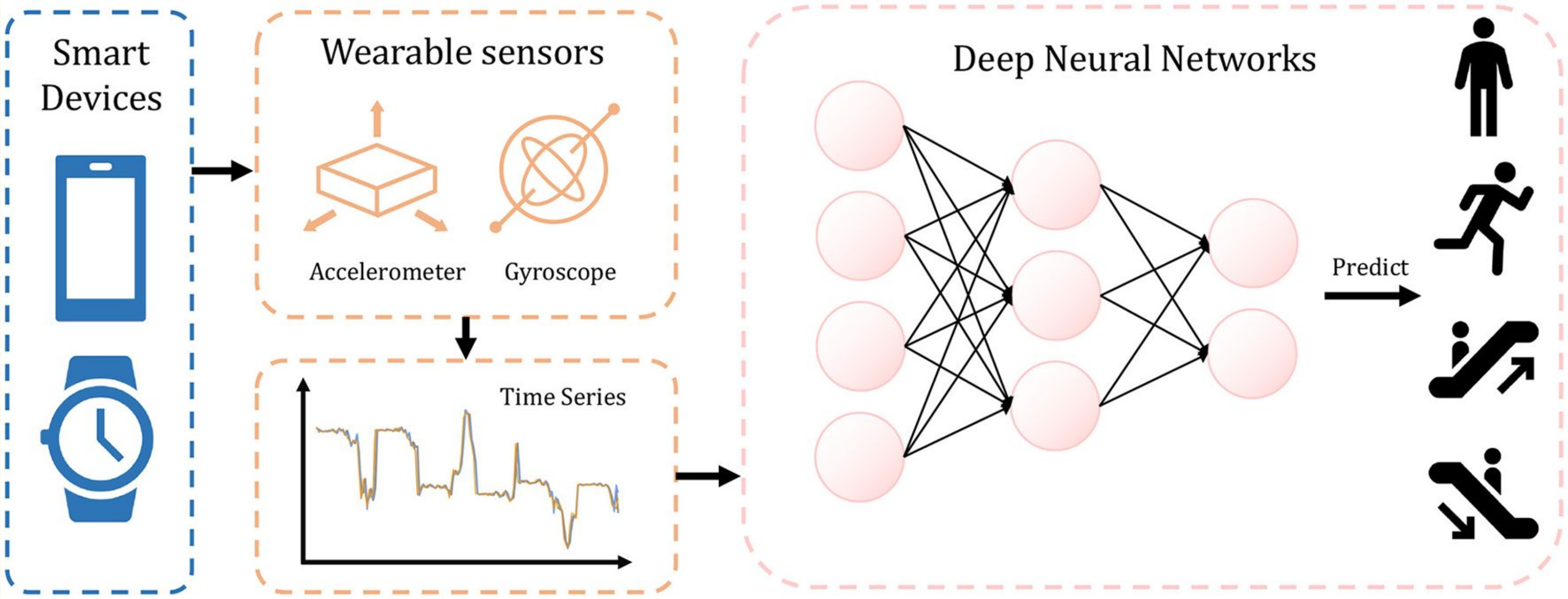
- This is **expensive**, **impractical**, and **difficult to scale**.

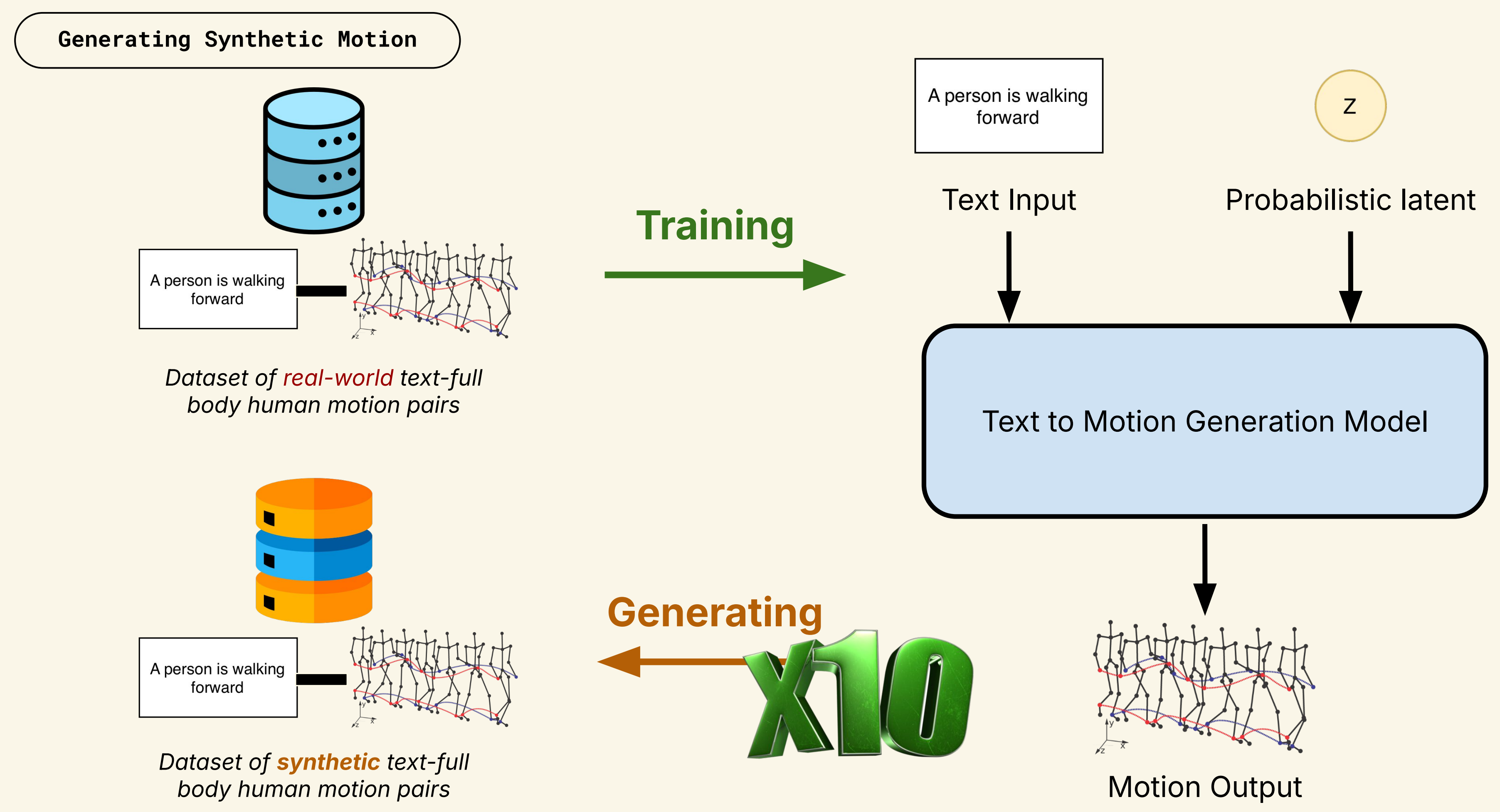
- Machine Learning runs on data (lots of data!) and needs **diverse**, **large** labelled datasets.

What if we use *synthetic data*?

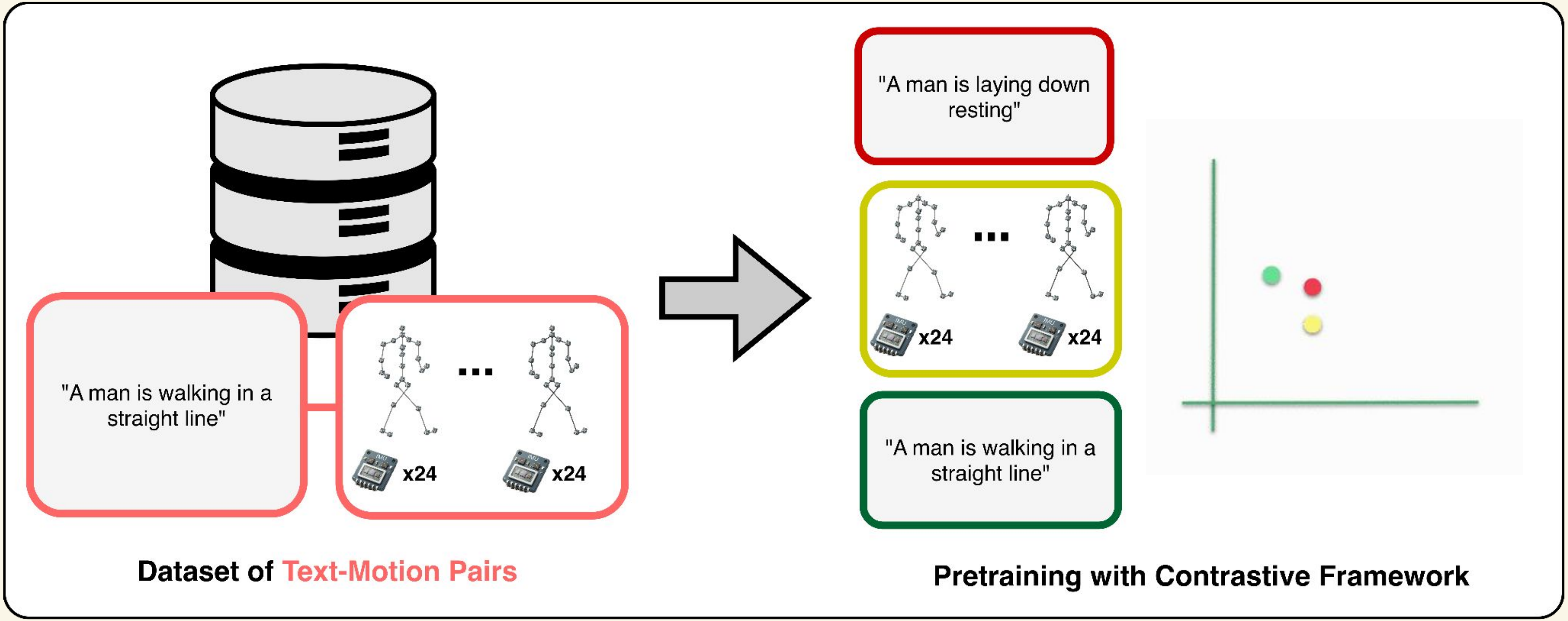
Wearables

Human Activity Recognition:




Generating Synthetic Motion
A person is walking forward
Dataset of real-world text-full body human motion pairs
Training
A person is walking forward
Text Input
Z
Probabilistic latent
Text to Motion Generation Model
Motion Output
Generating
X10
A person is walking forward
Dataset of synthetic text-full body human motion pairs

Pretraining Motion Time-Series Model



Pretraining with Synthetic Data

Table 2: Scaling synthetic data improves downstream HAR performance. At sufficient scale (8×), synthetic pretraining outperforms real-world pretraining in all fine-tuned settings, with the exception of the 0-shot regime. Reported scores are mean macro-averaged F1 across datasets; **bold** and <u>underlined</u> values denote best and second-best results per column, respectively. All results are obtained using a fixed random seed across runs.

Method	0-shot	1-shot	2-shot	3-shot	5-shot	10-shot
Real-World Baseline	**0.3070**	0.5445	0.6243	0.6611	0.7041	0.7667
MDM (Synthetic)						
1×	0.2460	0.5120	0.6160	0.6559	0.6974	0.7653
2×	<u>0.2928</u>	0.5334	0.6089	0.6505	0.7008	0.7368
4×	0.2546	0.5299	0.6344	0.6749	0.7023	0.7663
8×	0.2654	**0.5517**	0.6266	0.6704	0.7165	0.7792
T2M (Synthetic)						
1×	0.2606	0.5229	0.6100	0.6551	0.7024	0.7649
2×	0.2462	0.5132	0.6062	0.6669	**0.7226**	0.7635
4×	0.2482	0.5256	0.6114	<u>0.6766</u>	0.7024	0.7719
8×	0.2123	<u>0.5491</u>	<u>0.6372</u>	0.6631	0.7050	0.7891
MotionDiffuse (Synthetic)						
1×	0.2738	0.5363	0.6060	0.6548	0.7053	0.7739
2×	0.2739	0.5233	0.6186	0.6725	0.6995	0.7907
4×	0.2566	0.5423	0.6472	0.6700	0.7150	**0.7959**
8×	0.2427	0.5460	**0.6550**	**0.6893**	<u>0.7213</u>	<u>0.7956</u>

- **Consistent gap** remains in the 0-shot setting!

- Suggests that real-world data induces representations that generalise more effectively **out-of-the-box**, whereas synthetic pretraining excels under **adaptation**.

- Performance improves as synthetic data scales but **trend is not strictly monotonic**

Thank you!

Motion Capture is Not the Target Domain: Scaling Synthetic Data for Learning Motion Representations

Firas Darwish[*]
Department of Statistics
University of Oxford
Oxford, UK
{firas.darwish}@worc.ox.ac.uk

George Nicholson
Department of Statistics
Big Data Institute
University of Oxford
Oxford, UK
{george.nicholson}@stats.ox.ac.uk

Aiden Doherty[†]
Nuffield Department of Population Health
Big Data Institute
University of Oxford
Oxford, UK
{aiden.doherty}@ndph.ox.ac.uk

Hang Yuan[†]
Nuffield Department of Population Health
Big Data Institute
University of Oxford
Oxford, UK
{hang.yuan}@ndph.ox.ac.uk

Abstract

Synthetic data offers a compelling path to scalable pretraining when real-world data is scarce, but models pretrained on synthetic data often fail to transfer reliably to deployment settings. We study this problem in full-body human motion, where large-scale data collection is infeasible but essential for wearable-based Human Activity Recognition (HAR), and where synthetic motion can be generated from motion-capture–derived representations. We pretrain motion time-series models using such synthetic data and evaluate their transfer across diverse downstream HAR tasks. Our results show that synthetic pretraining improves generalisation when mixed with real data or scaled sufficiently. We also demonstrate that large-scale motion-capture pretraining yields only marginal gains due to domain mismatch with wearable signals, clarifying key sim-to-real challenges and the limits and opportunities of synthetic motion data for transferable HAR representations.