ISSAI

NAZARBAYEV UNIVERSITY | Institute of Smart Systems and Artificial Intelligence

Deep Learning as an Assistive Technology for Cognitive and Visual Impairments

Prof. Huseyin Atakan Varol
Founding Director

ahvarol@nu.edu.kz
issai.nu.edu.kz

AIPPA 2023 | Artificial Intelligence - Possibilities for Practical Applications
Искусственный интеллект - возможности практического применения

Institute of Smart Systems and Artificial Intelligence (ISSAI) was established in September 2019:

- to develop a national capacity for research in AI on the basis of Nazarbayev University
- as a leading AI institute in Kazakhstan
- to conduct cutting-edge research applicable not only for Kazakhstan but also for the world
- incorporating the experience of exemplars from Asia, Europe and the North America
- to provide an agile framework for research, innovation and collaboration with national and international partners in education, industry and government
- following highest standards for ethics and integrity (AI for Good)

ISSAI
NAZARBAYEV UNIVERSITY | Institute of Smart Systems and Artificial Intelligence

ISSAI team consists of over 30 researchers, 30 data curators, and 5 affiliated faculty members.

ISSAI Computational Resources for AI

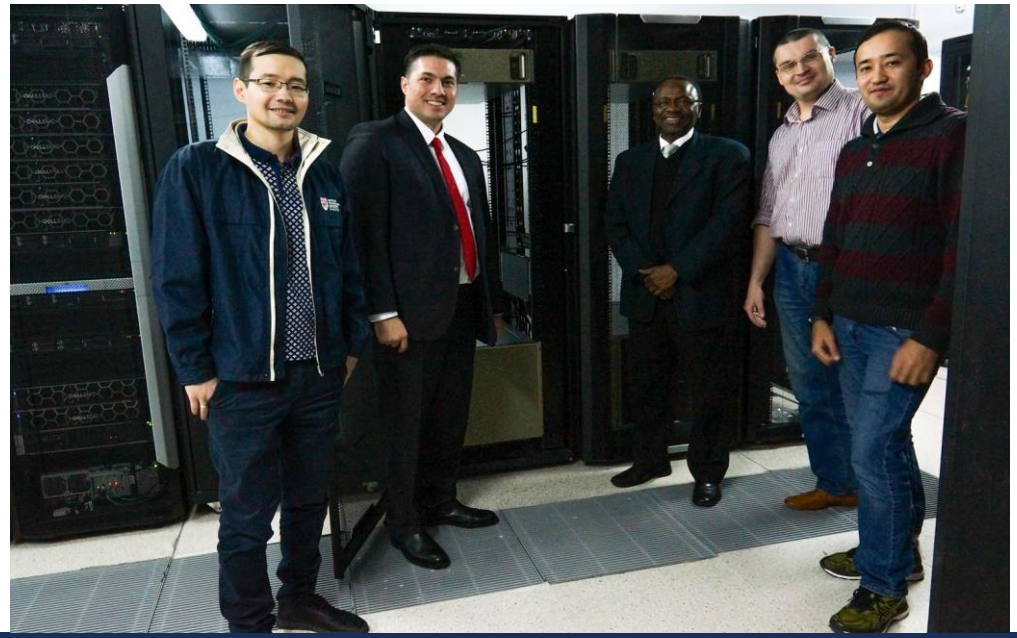
ISSAI has established the most advanced AI computational lab in the country, as AI requires enormous datasets and computational resources for training the models.

DGX supercomputers from Nvidia are the de-facto standard for AI research.

In Kazakhstan, ISSAI was first to operationalize the DGX systems and has garnered substantial expertise in the use of them.

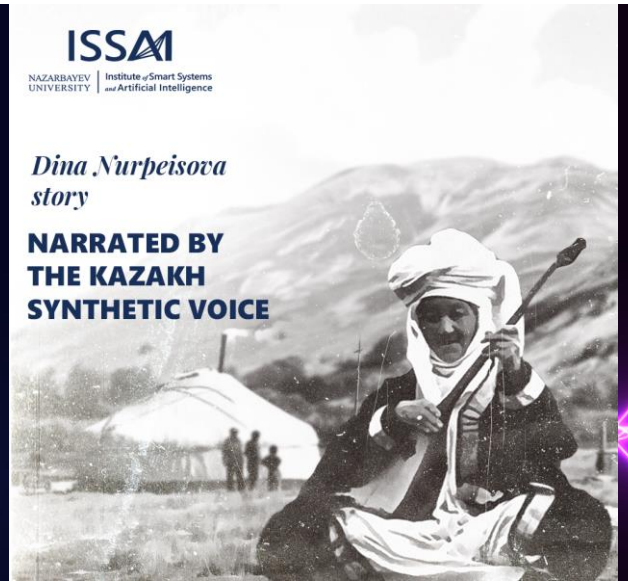
- 2 DGX-1 units in 2019
- 2 DGX-2 units in 2020
- 4 DGX-A100 units in 2022

ISSAI shares its computational resources with NU researchers and other Kazakhstani universities.



ISSAI Roadmap for AI-Based Speech and Natural Language Processing Technologies

KAZAKH TEXT-TO-SPEECH CONVERSION TECHNOLOGY



ISSAI is contributing to the soft digital infrastructure of Kazakhstan
by creating digital public goods that become vital components of AI-based products and services.

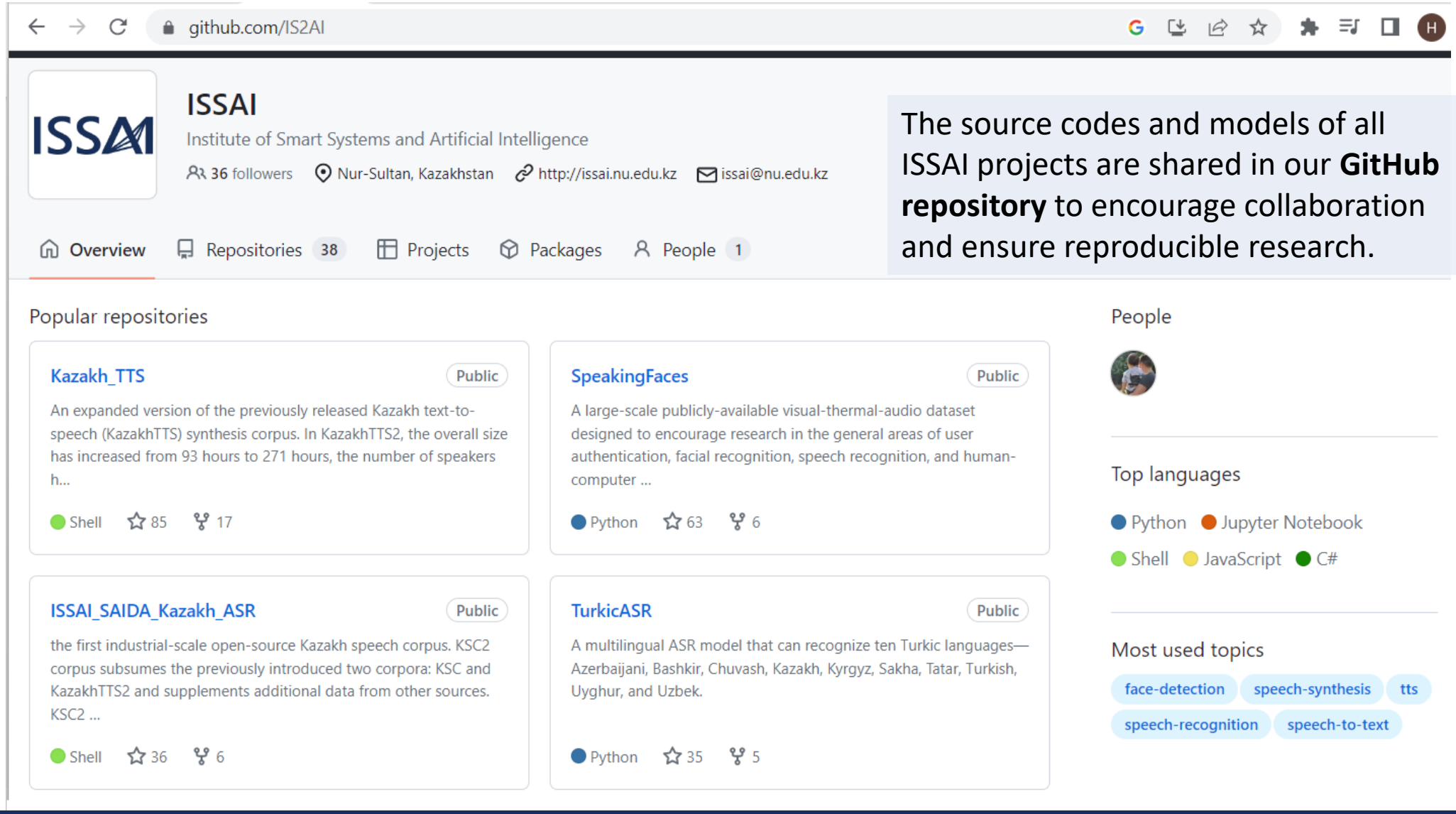
ISSAI as the Largest Repository of Open-Source AI Datasets in Central Asia

Unique to ISSAI in Central Asia, all research datasets created by the Institute are open-source and publicly available for download from ISSAI's website.

DataCite

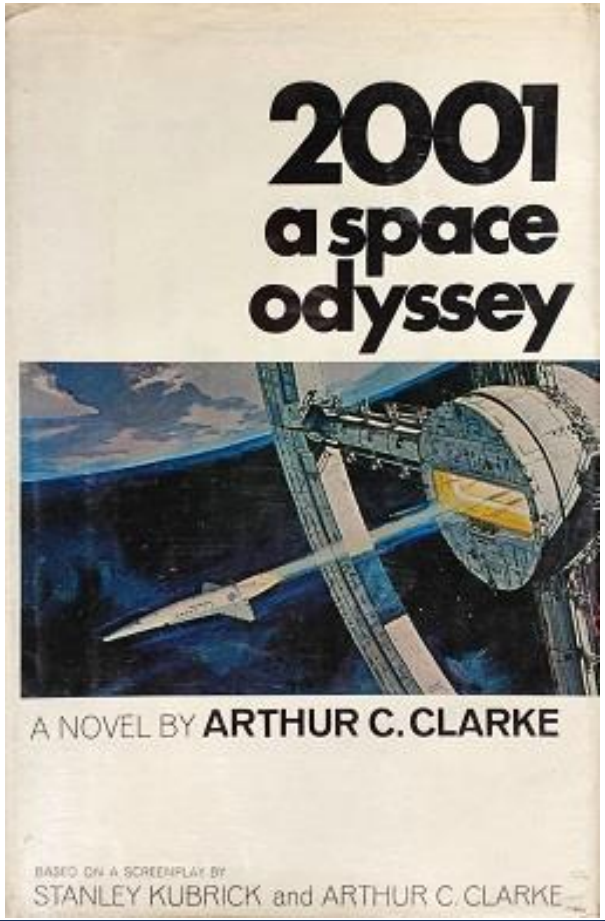
- Kazakh Speech Corpus (KSC)
- Russian Speech Corpus (OpenSTT-CS334)
- WiFi Signature and IMU-based Indoor Localization (IMUWiFine)
- Kazakh Text-to-Speech (KazakhTTS)
- Uzbek Speech Corpus (USC)
- Thermal and Visual Images Synchronized with Audio (SpeakingFaces)
- Kazakh-accented English (K-Eng)
- WiFi Signature-based Indoor Localization (WiFine)
- Thermal Faces in the Wild (TFW)

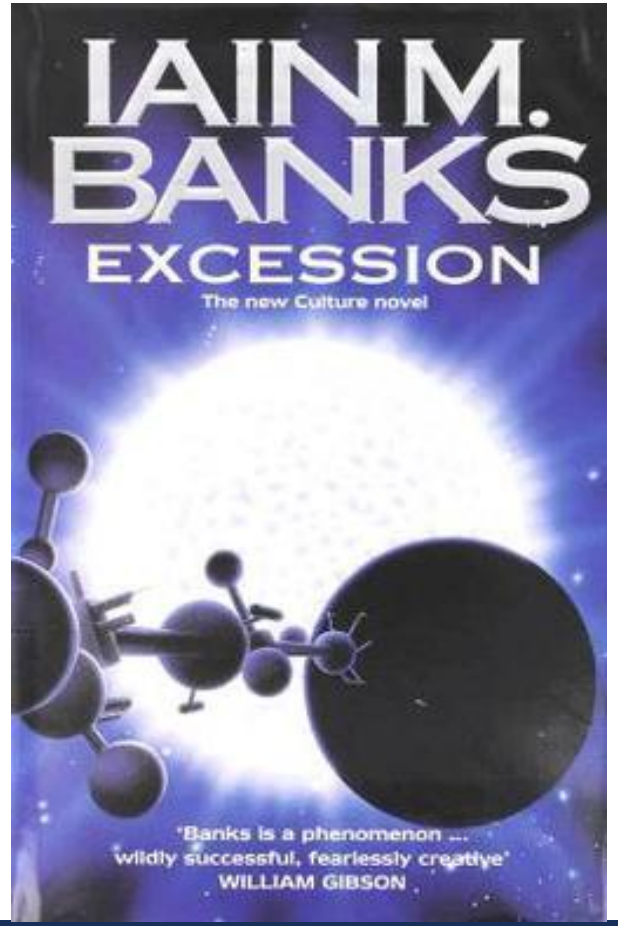
ISSAI as a Proponent of Reproducible AI Research

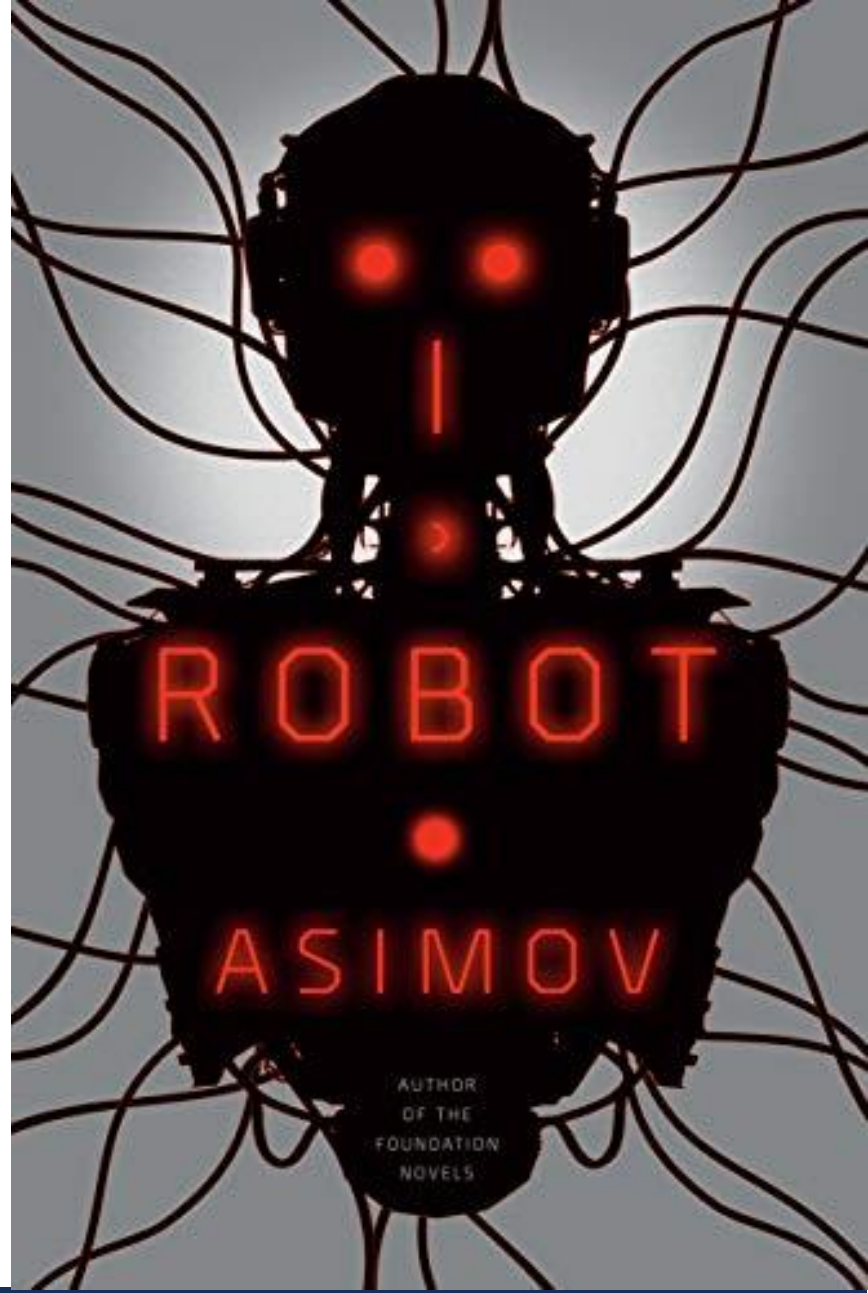


The source codes and models of all ISSAI projects are shared in our **GitHub repository** to encourage collaboration and ensure reproducible research.

The **technological singularity** is a hypothetical future point in time at which technological growth becomes uncontrollable and irreversible, resulting in unforeseeable changes to human civilization.







What if we already reached technological singularity in 2023?

30 November 2022 – Launch of ChatGPT

ChatGPT is a conversational agent developed by OpenAI.

It is trained on a huge text corpus using a dedicated data center with AI servers.

It is based on the GPT-3 large scale language model.

It was fine-tuned using human in the loop reinforcement learning.

It can generate realistic text in any domain.

It is one of the first attempts for **artificial general intelligence (AGI)**.

Artificial Narrow Intelligence (ANI)

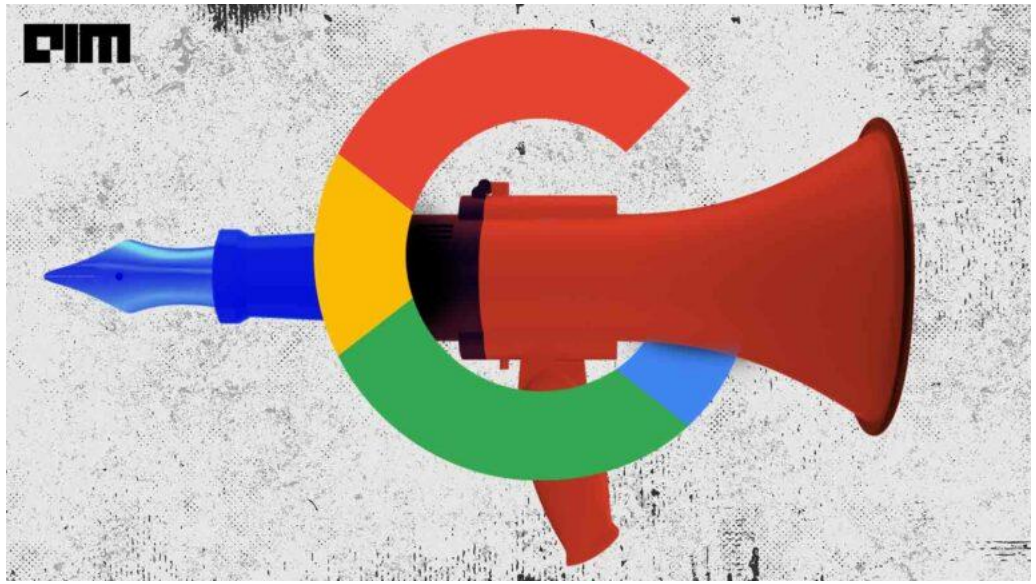
ANI can perform a single task (e.g., translation, food recognition, spam filtering)
ANI cannot carry out tasks outside its designed function.
ANI outperforms humans in all narrow task domains.

Artificial General Intelligence (AGI)

A wide set of human capabilities often surpassing human performance.

6 March 2023 – Launch of Google Universal Speech Model (USM)

USM is a speech recognition model with 2 billion parameters trained on 12 million hours of speech and 28 billion sentences of text, spanning 300+ languages.



1 March 2023 – Launch of Microsoft Kosmos Language Vision Model

What's in this picture?

Looks like a duck.

That's not a duck. Then what's it?

Looks more like a bunny.

Why?

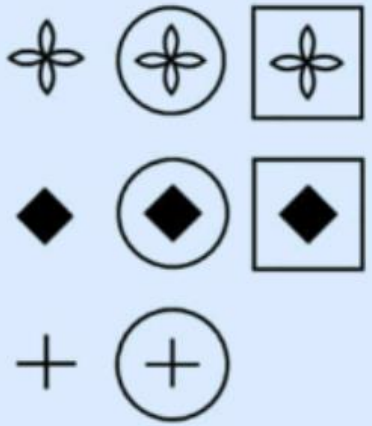
It has bunny ears.

Description of three toed woodpecker: It has black and white stripes throughout the body and a yellow crown.
Description of downy woodpecker: It has white spots on its black wings and some red on its crown.

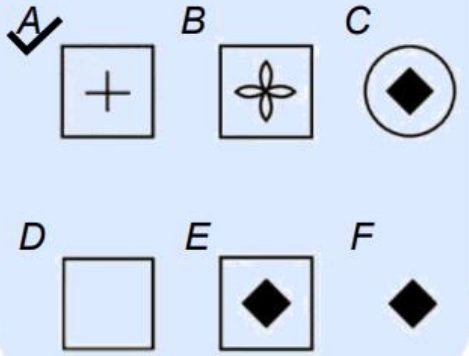
Question: what is the name of the woodpecker in the picture?

Downy

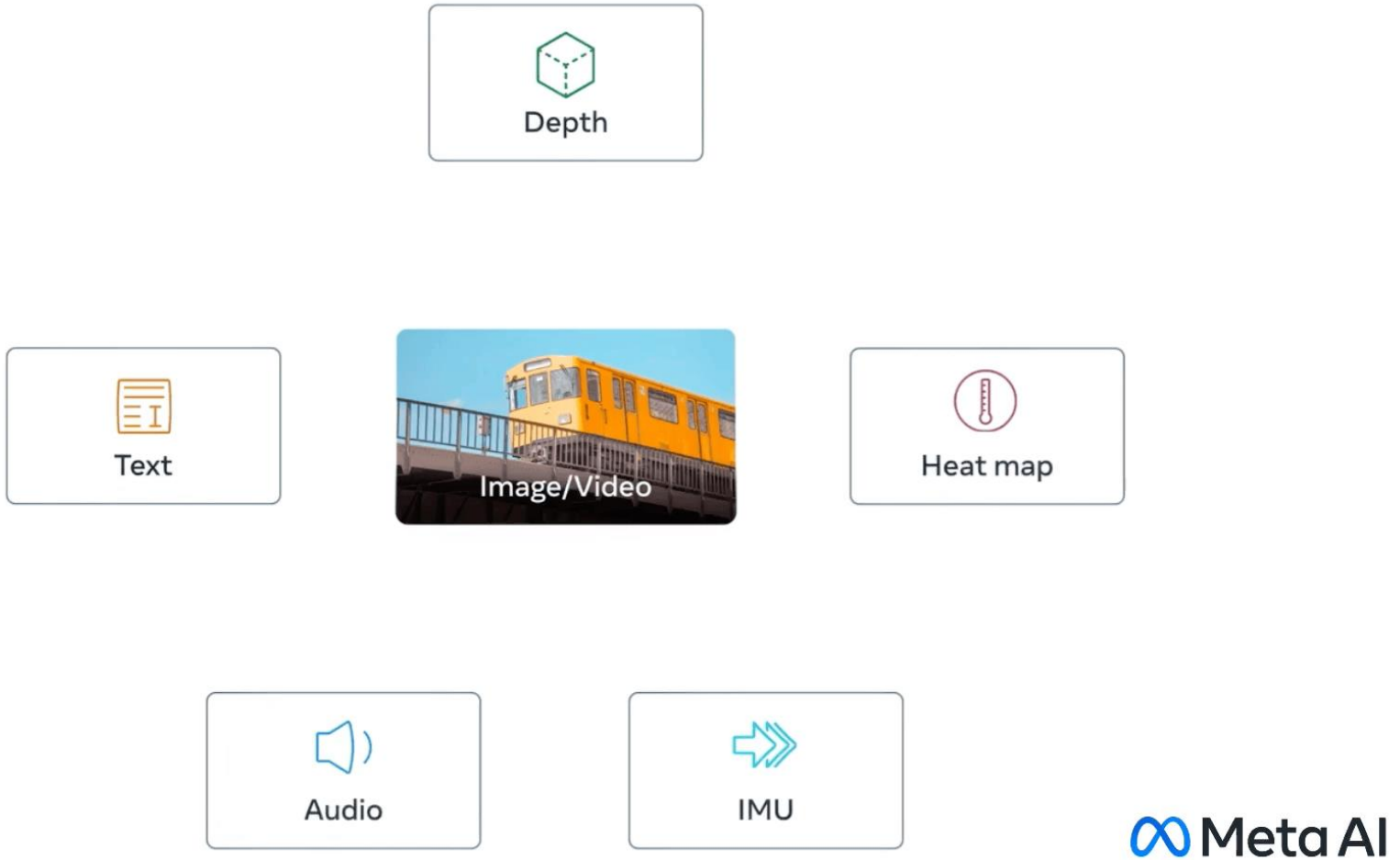
Here are eight images:



The following image is:

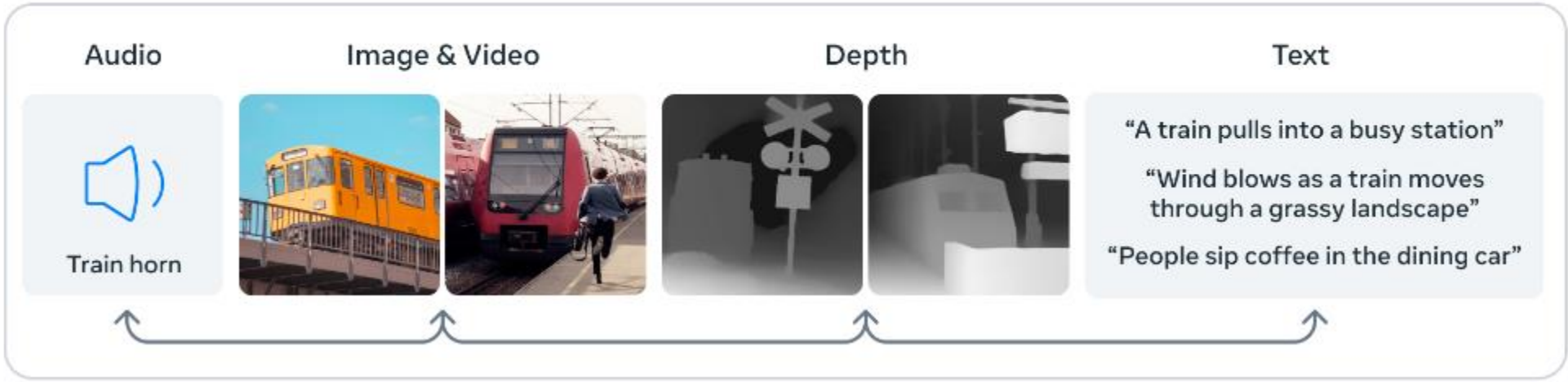


9 May 2023 – ImageBind: Holistic AI learning across six modalities



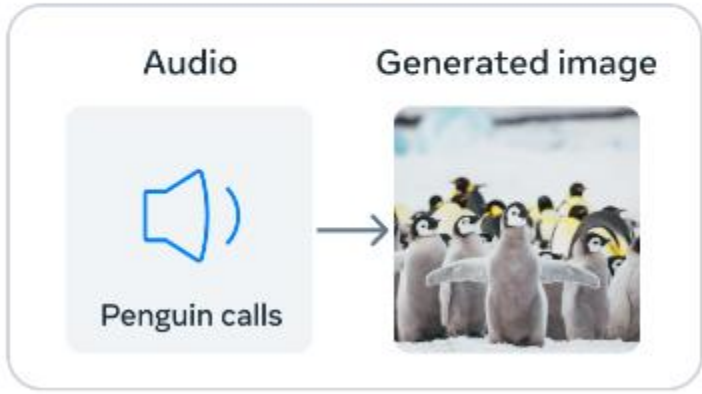
9 May 2023 – ImageBind: Holistic AI learning across six modalities

Cross-modal retrieval



Embedding-space arithmetic

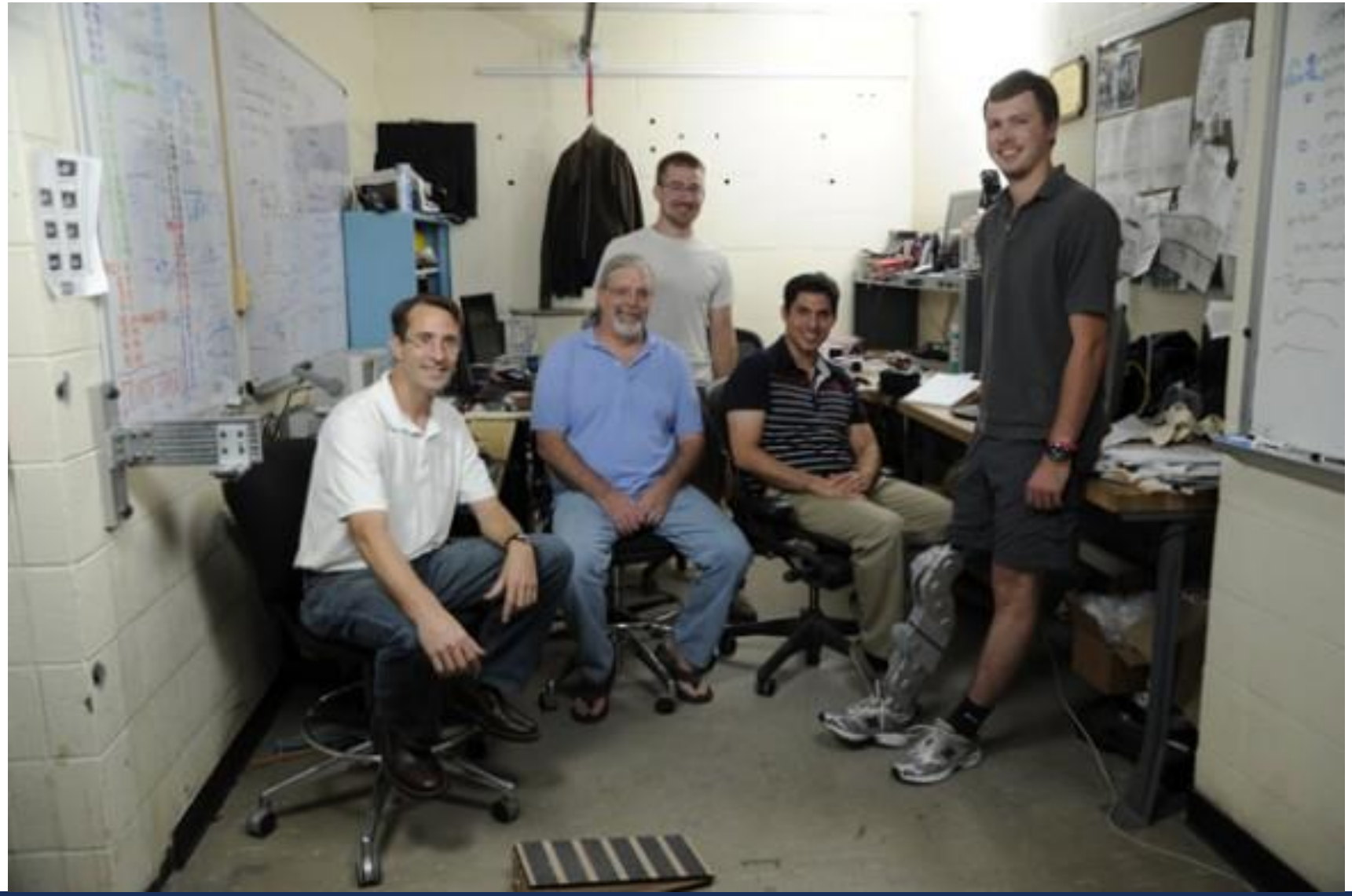
Audio to image generation



Assistive technology is an umbrella term covering the systems and services related to the delivery of assistive products and services.

Assistive products maintain or improve an individual's functioning and independence, thereby promoting their well-being.

My PhD focused on the use artificial intelligence in the control of robotic prostheses to restore near-normal function for physical disabilities.

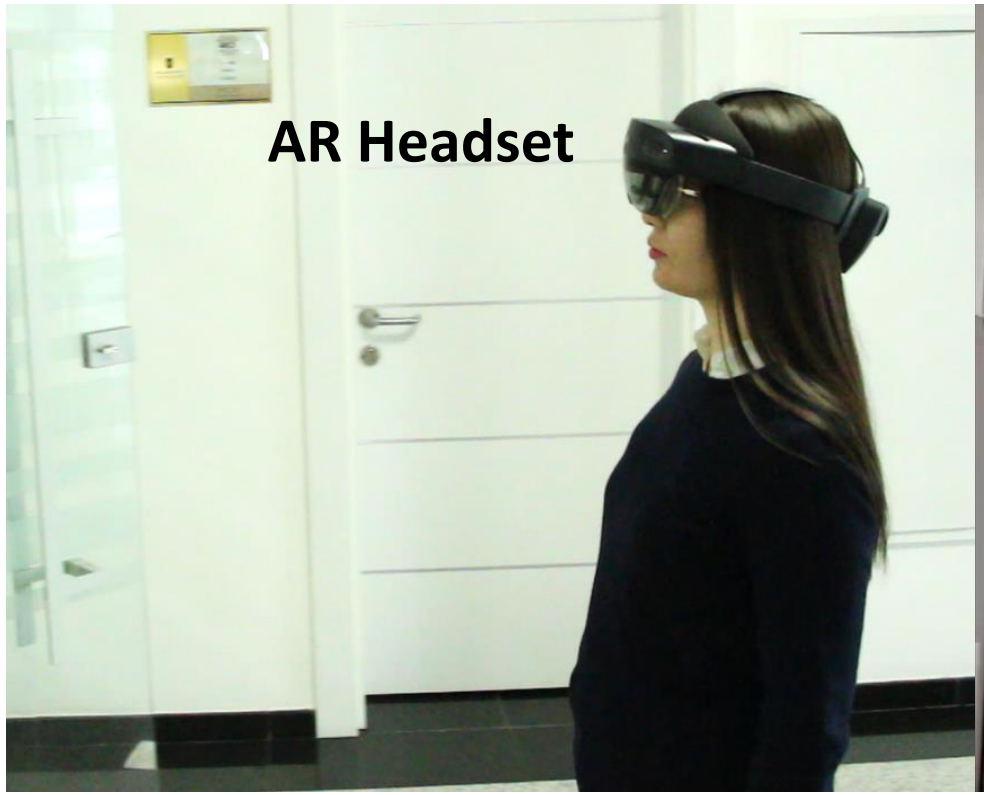


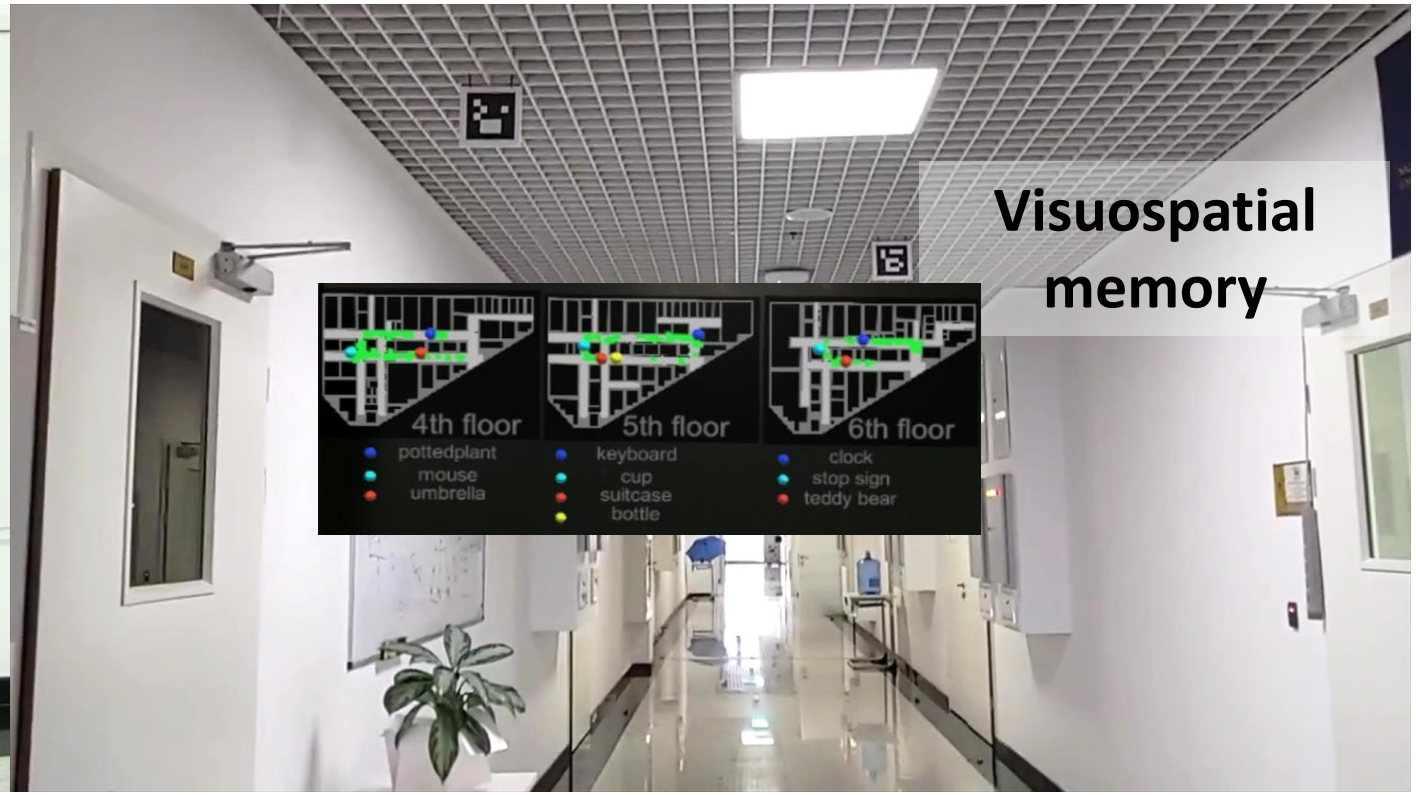
H. A. Varol, F. Sup and M. Goldfarb, "Multiclass Real-Time Intent Recognition of a Powered Lower Limb Prosthesis," in *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 542-551, March 2010,

Can we leverage advances in AI to create assistive devices for cognitively and visually impaired?

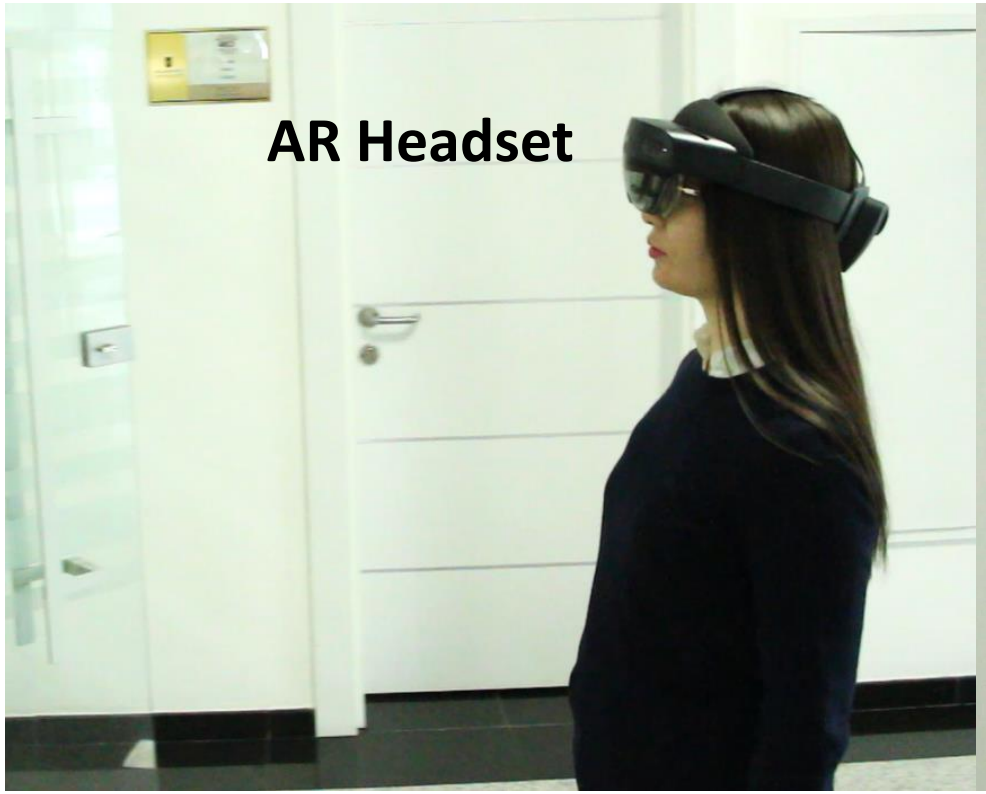
H. A. Varol, F. Sup and M. Goldfarb, "Multiclass Real-Time Intent Recognition of a Powered Lower Limb Prosthesis," in *IEEE Transactions on Biomedical Engineering*, vol. 57, no. 3, pp. 542-551, March 2010,

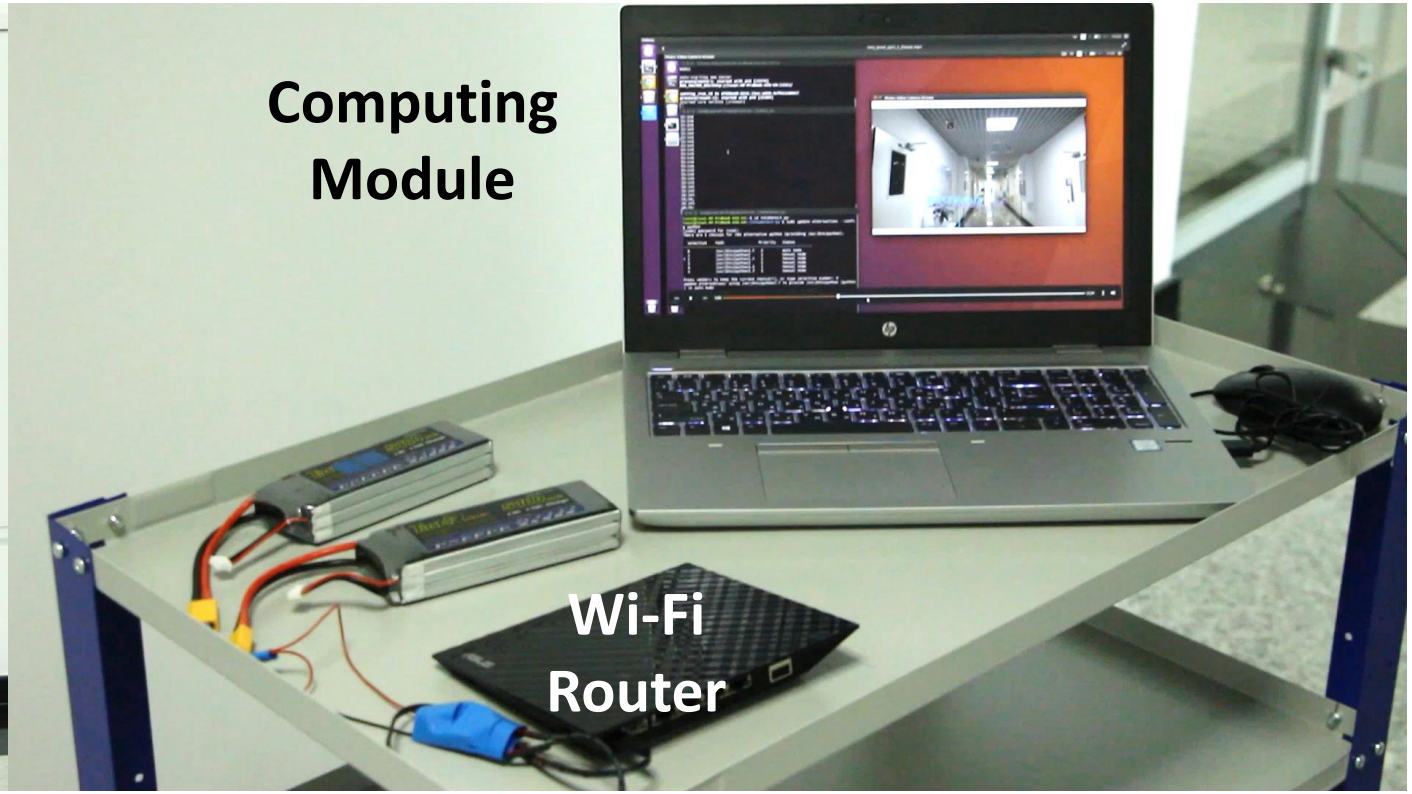
First **human memory** augmentation system that can construct a **holographic visuospatial memory** in an indoor environment, **ExoMem**.





Our solution leverages **augmented reality (AR)** and **artificial intelligence (AI)** and comprises two components:





Augmented Reality (AR)

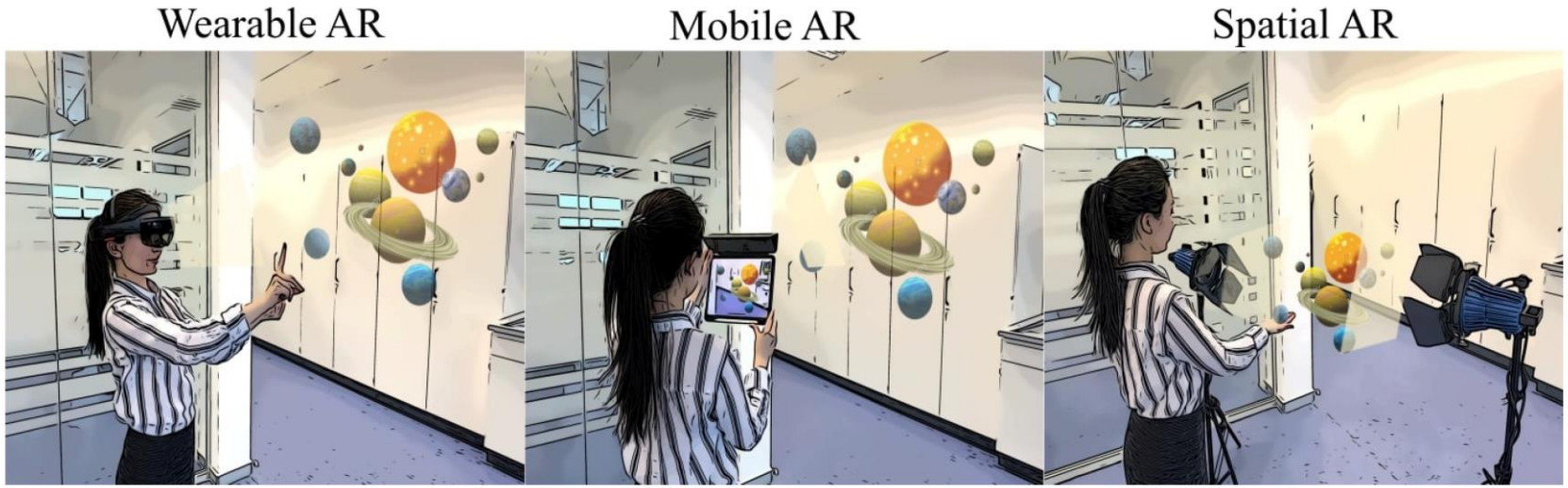
Augmented Reality (AR) is an immersive experience that superimposes virtual 3D objects upon a user's direct view of the surrounding real environment, generating the illusion that those virtual objects exist in that space.

1960 - "**Man-Computer Symbiosis**", J. C. R. Licklider, MIT

1968 - **"A research center for augmenting human intellect"**, D. Engelbart, Stanford

1993 - **"Thinking with machines: intelligence augmentation"**, P. Skagestad **,**University of Massachusetts Lowell

First VR Headset "The ultimate display"
Ivan Sutherland, 1965, MIT Lincoln Lab.



AR Goggles as Sensing, Computation and Visualization Platforms

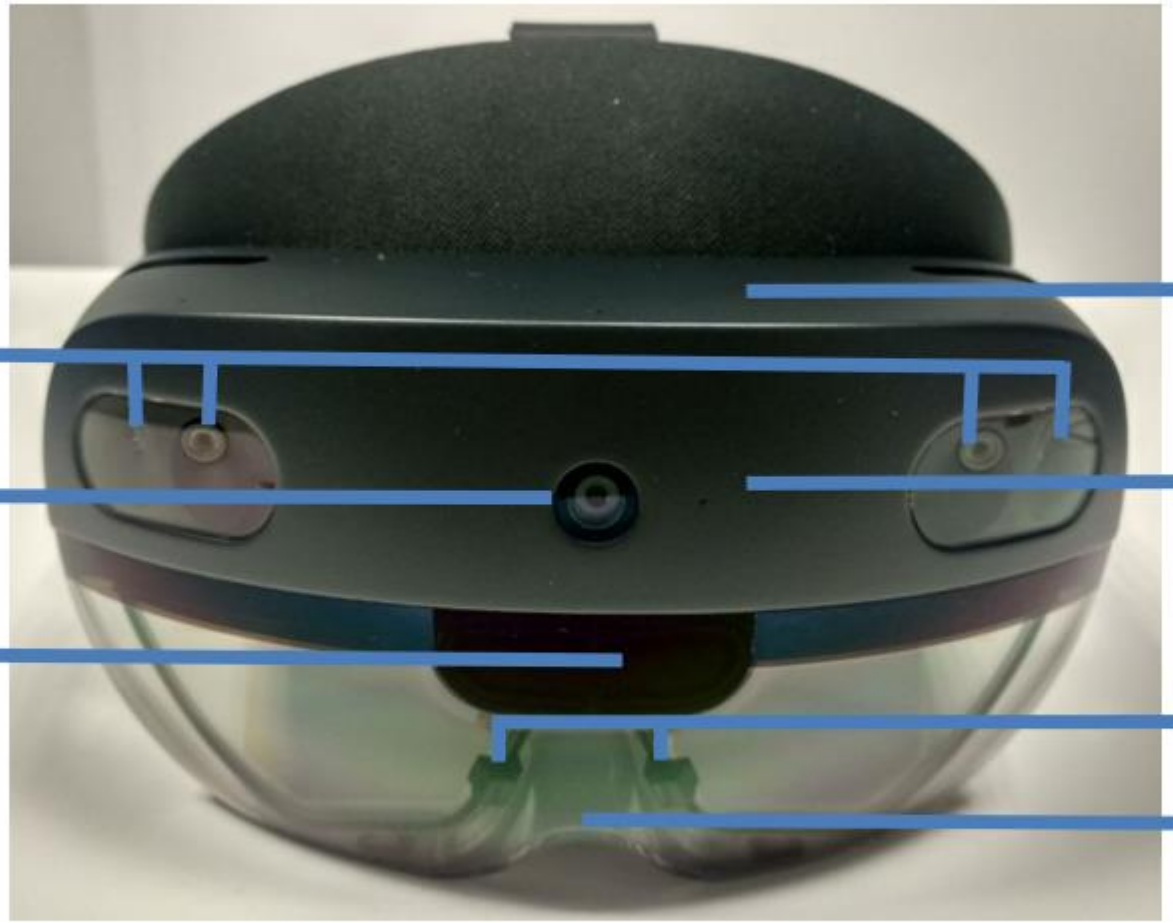
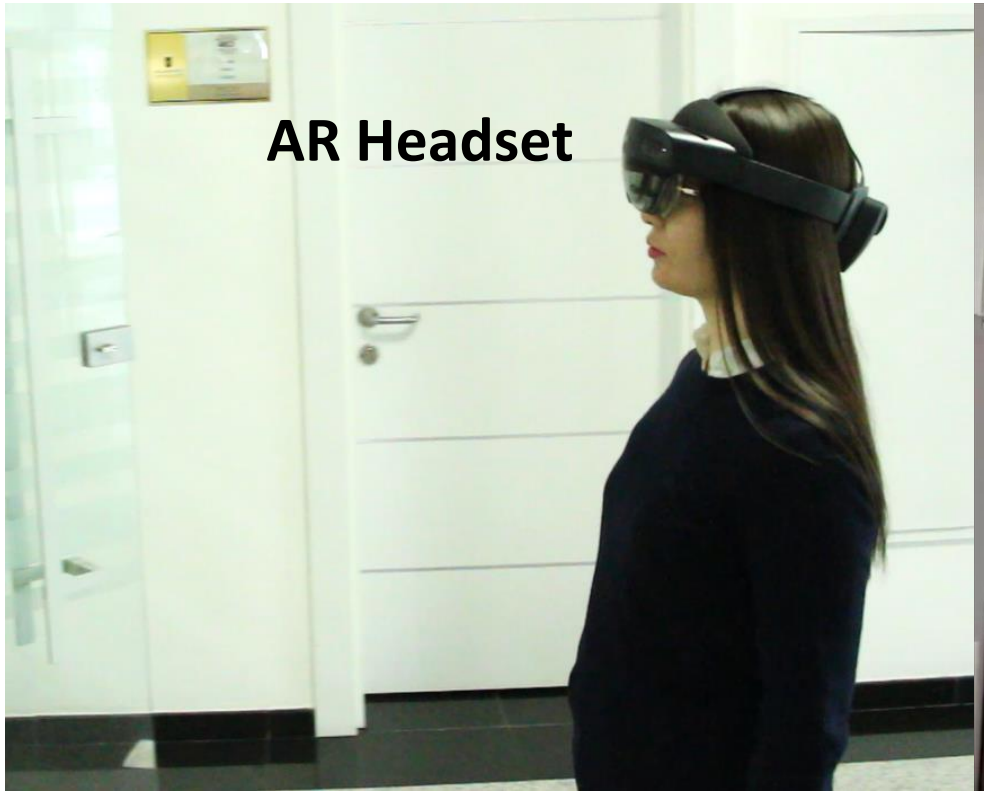
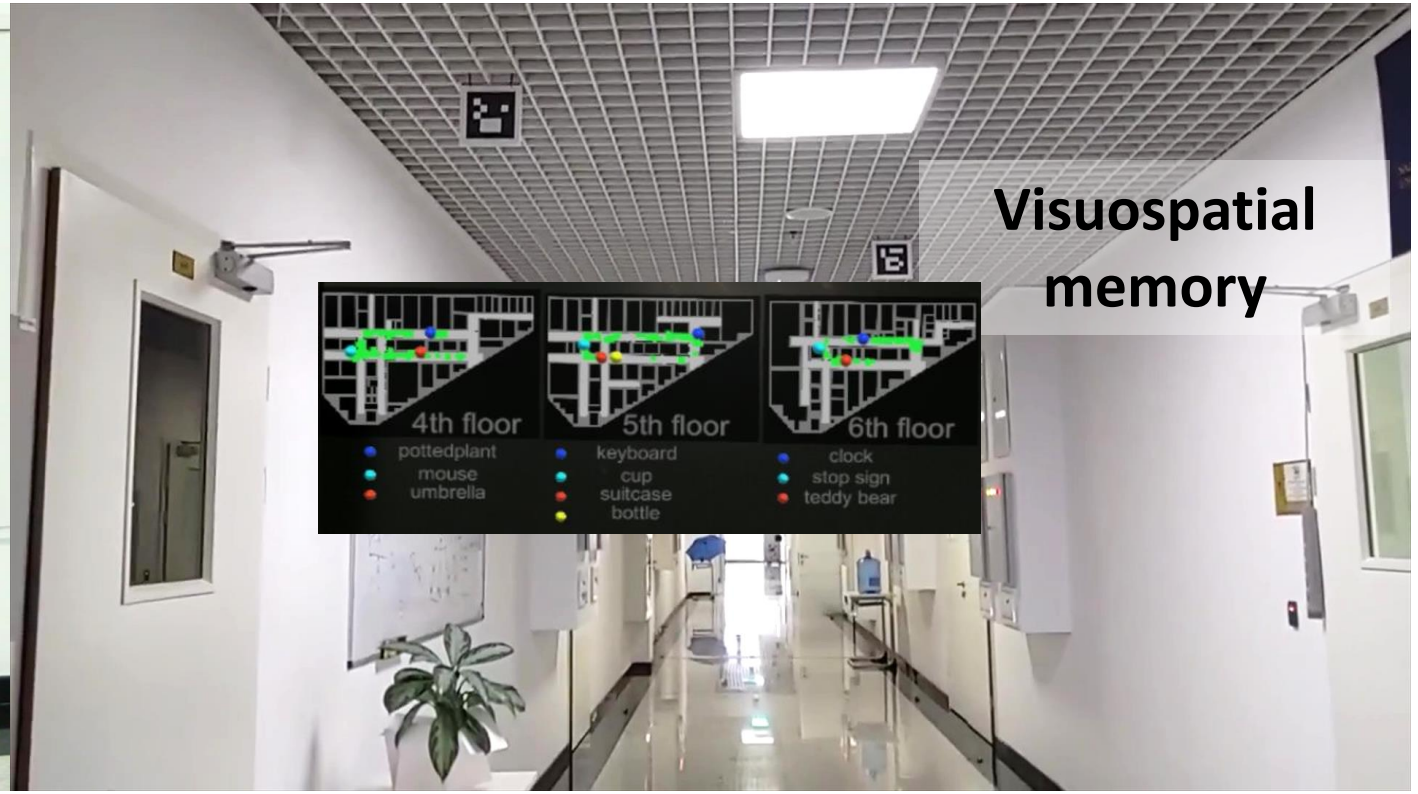


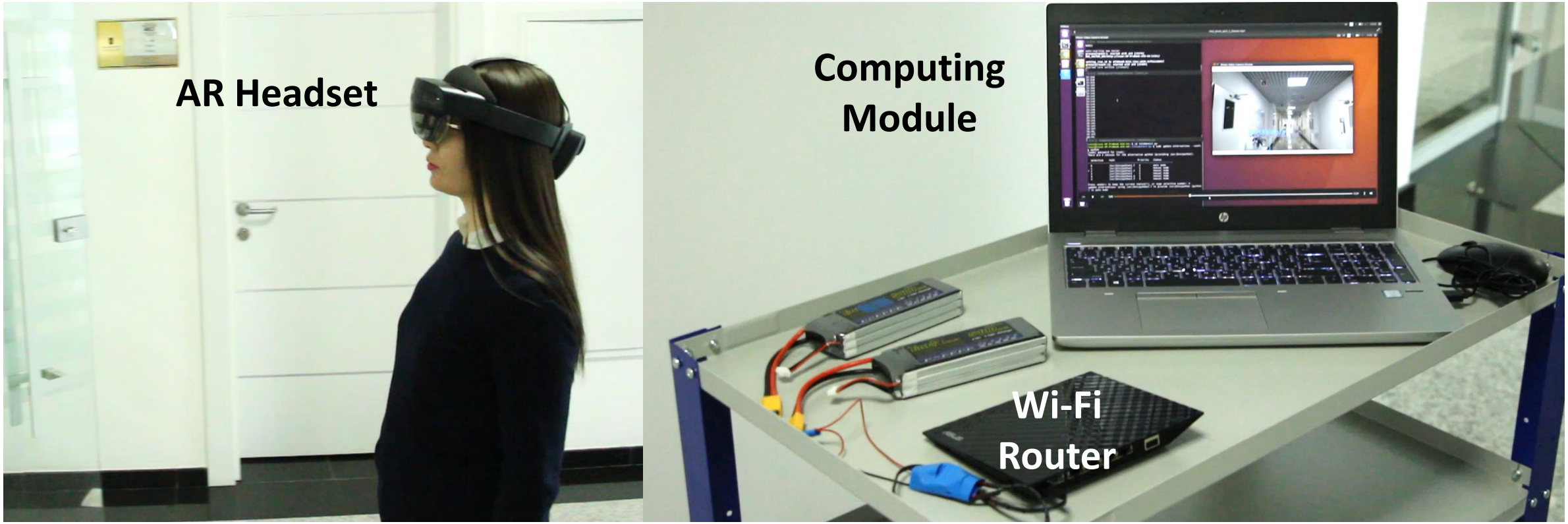
Fig 1. Microsoft Hololens 2 AR Goggles

1. **AR headset** that senses the environment, exchanges data over a wireless network and constructs a **holographic visuospatial memory** in the AR environment.

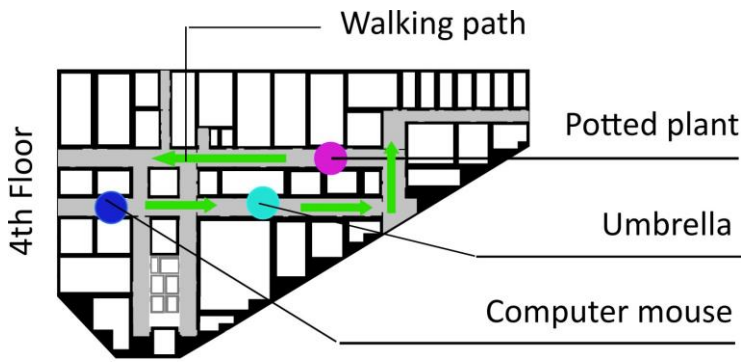


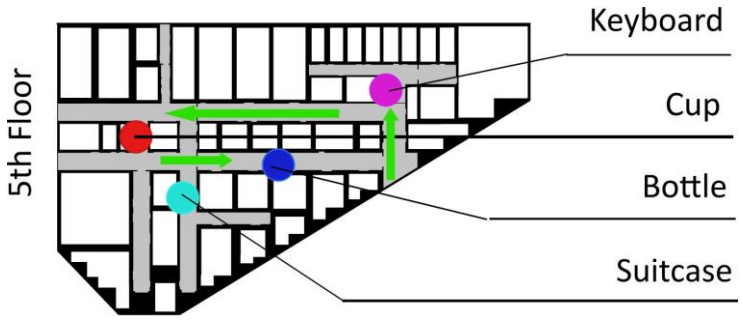


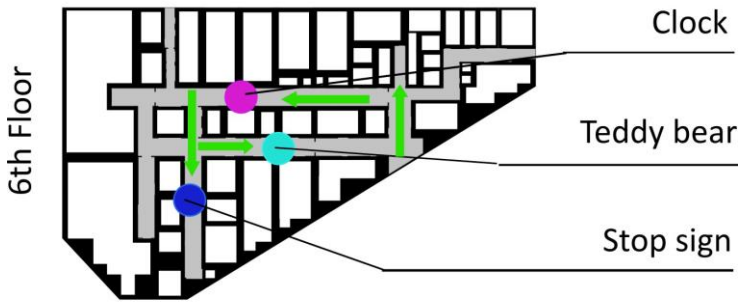
2. **A computing module** that performs computer vision-based **localization** and **object detection** on first-person view (FPV) data received from the AR headset.



To validate our system, we designed **object location memory task** that involved **two activities** (i.e., object location memorization and map-pointing). We evaluated the **cognitive load** of people on experimental studies.







In the **object location memorization activity**, participants completed a 20-minute tour of the **three floors** of the building and had to memorize the location of **ten** different **objects** they saw along the path.

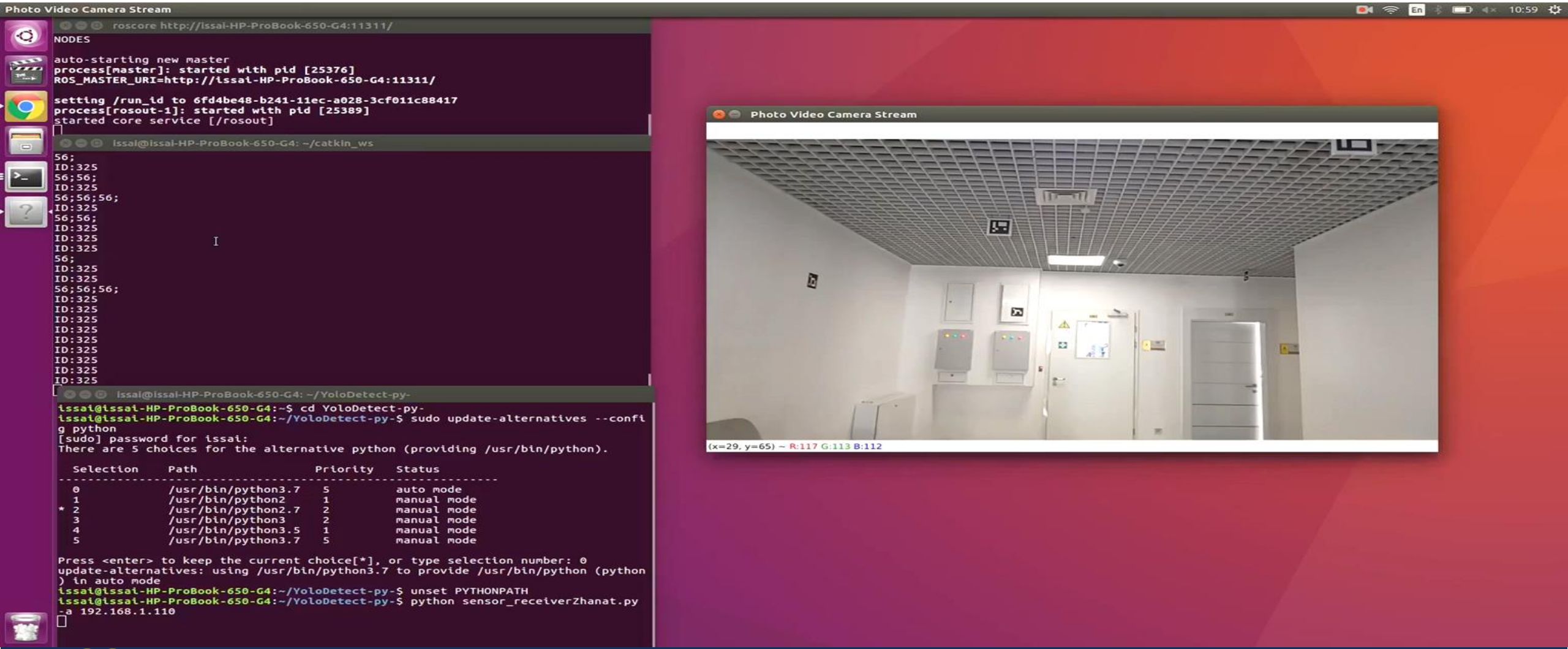
In the **map-pointing activity**, participants had to recognize seen objects and mark the object locations on the **map** presented in a **computer-based test** with and without ExoMem.

Participant completing the **object location memorization activity** with the ExoMem.

Third Person View

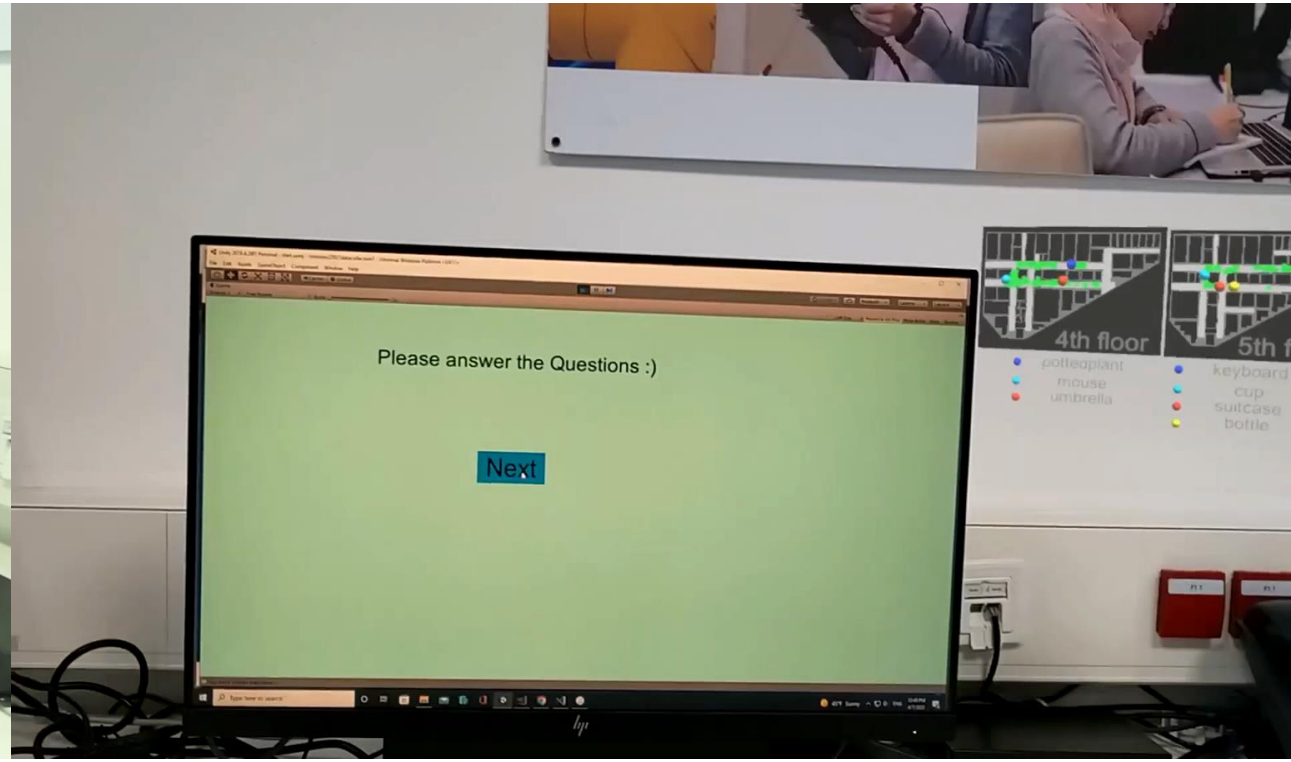
First Person View

CV-based user localization using **ArUco** fiducial markers and object recognition using **YOLO V4** object detector.



Participant completing the **map-pointing activity** with ExoMem.

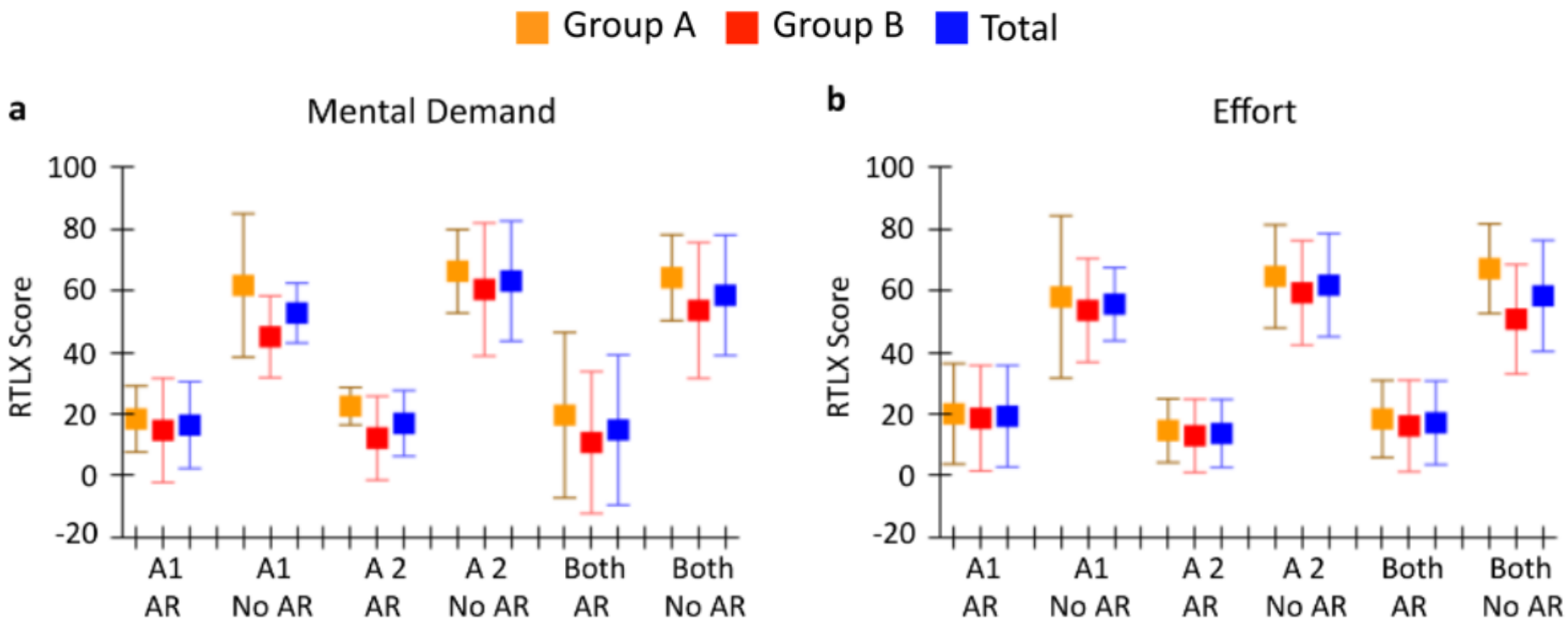
User completing the test



First person view with AR hologram

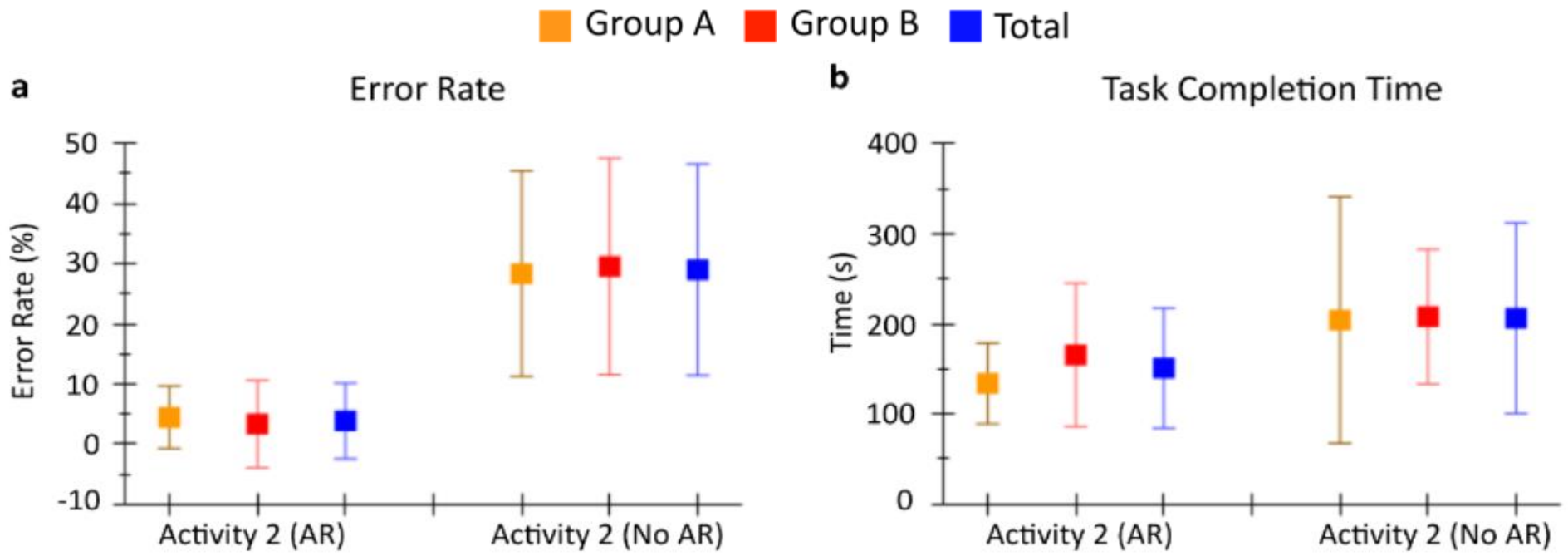
When using ExoMem, participants experienced much less **mental demand** and **effort** in both activities.

Nasa Task Load Index (NASA-TLX) Scores

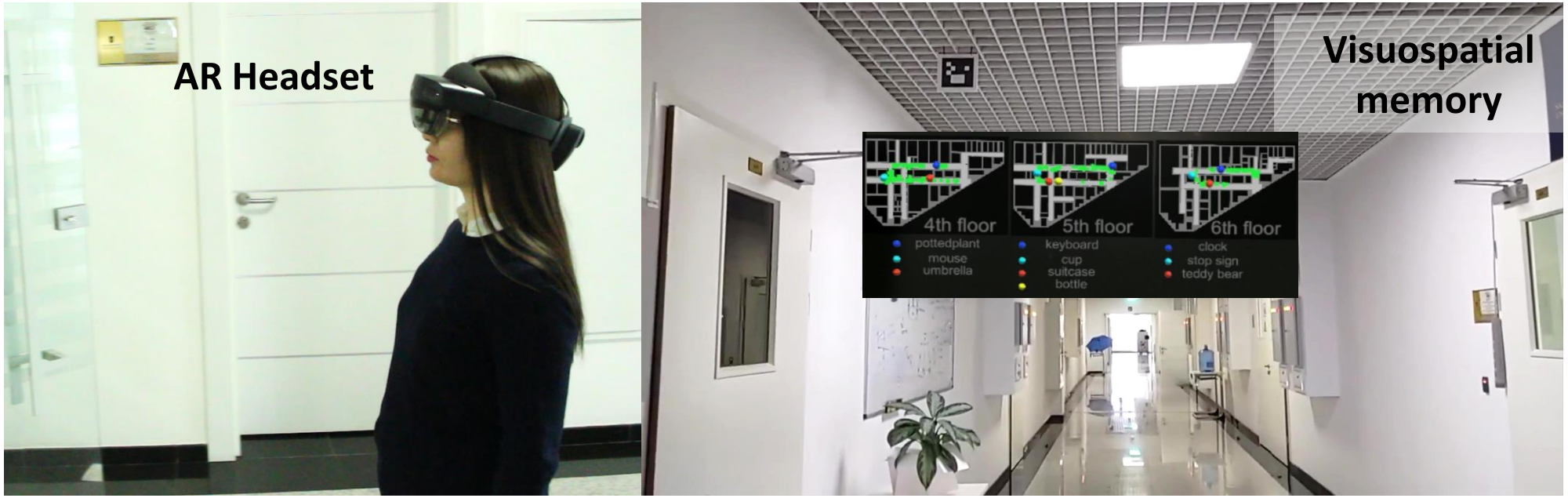


Objective evaluations indicated that participants made **7.52 times fewer errors** on the map-pointing activity and completed the computer-based test spending **27% less time** when using ExoMem.

Objective assessment of performance



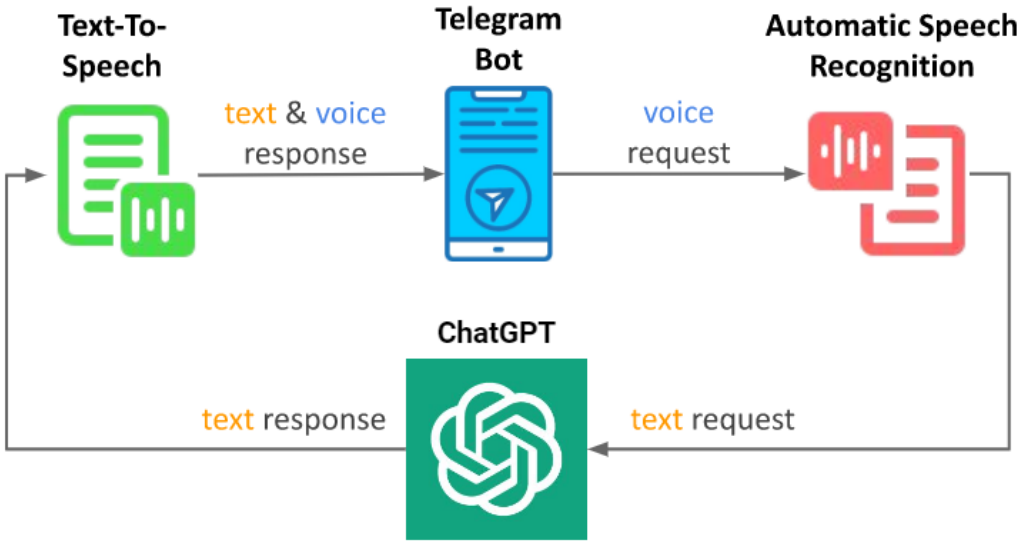
The results of our study highlight the potential of **AR** and **AI** technologies in developing **novel solutions** for generation, retrieval, and visualization of **visuospatial memory** for indoor environments.

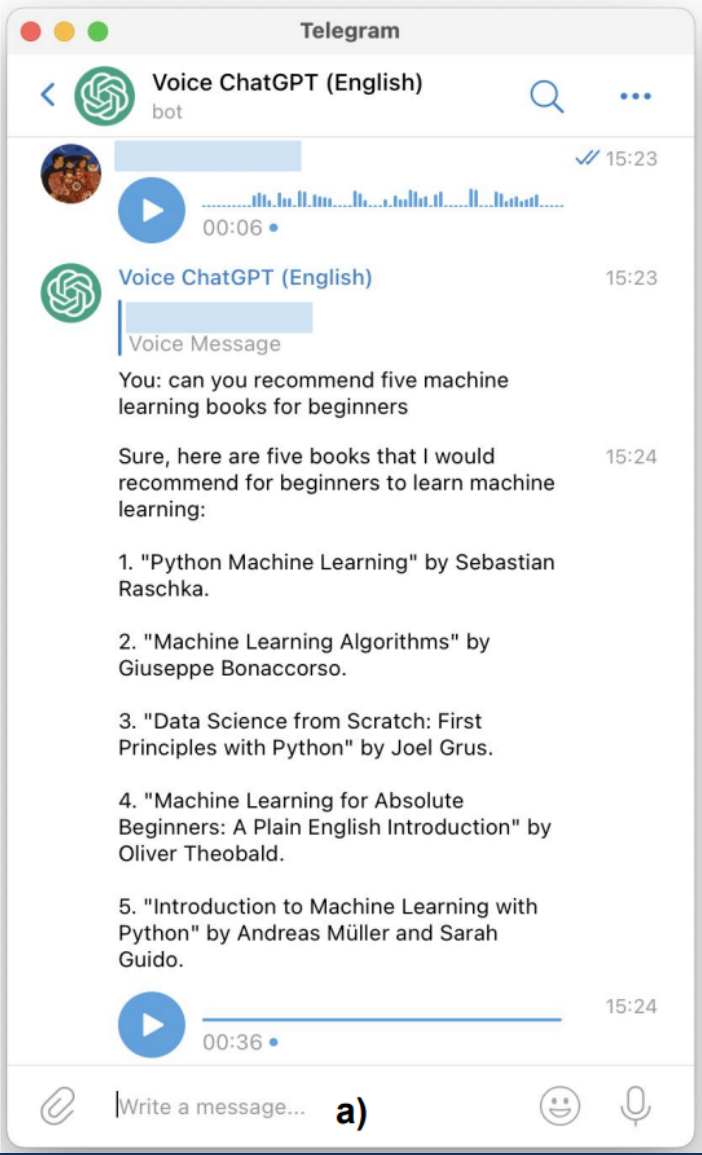


ChatGPT for Visually Impaired and Blind

ChatGPT is a large language model optimized for dialogues created by OpenAI.

In this project, we use ChatGPT, combined with automatic speech recognition and text-to-speech synthesis to serve as an assistive technology for accessing information for the blind and visually impaired.





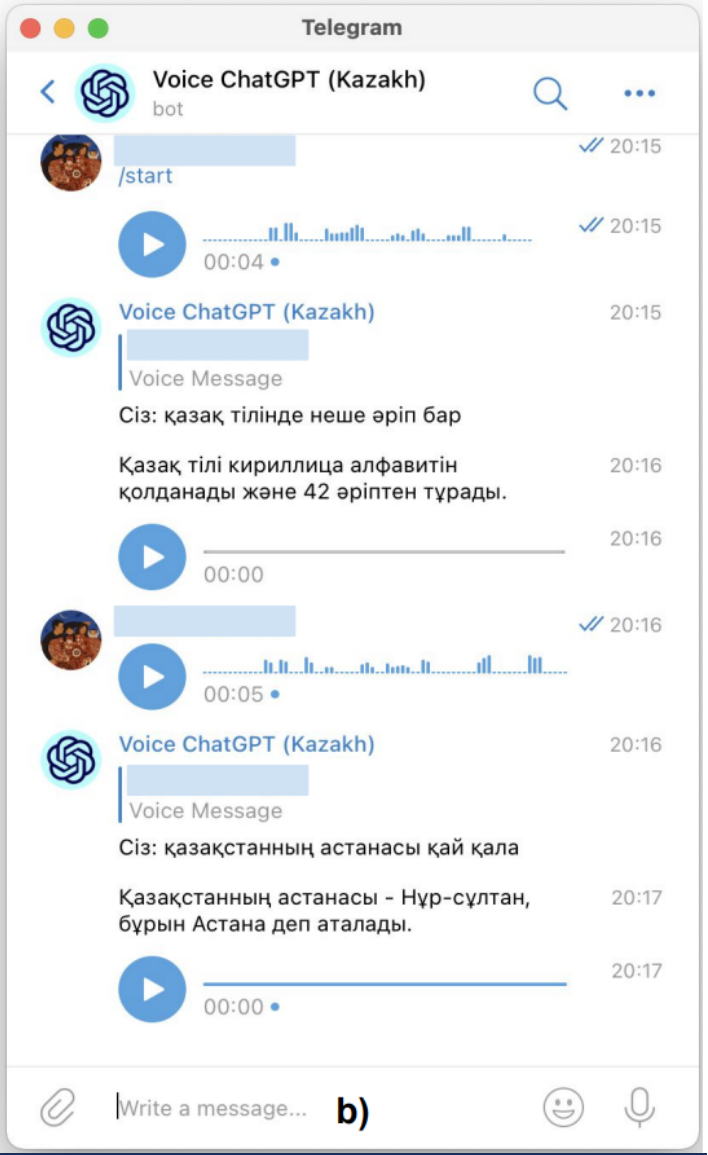


Image Captioning for the Visually Impaired and Blind

Visually impaired and blind people often face a range of socioeconomic problems that can make it difficult for them to live independently and participate fully in society.

Our system can provide the user with descriptive auditory feedback in the Kazakh language on a scene acquired in real-time by a head-mounted camera.

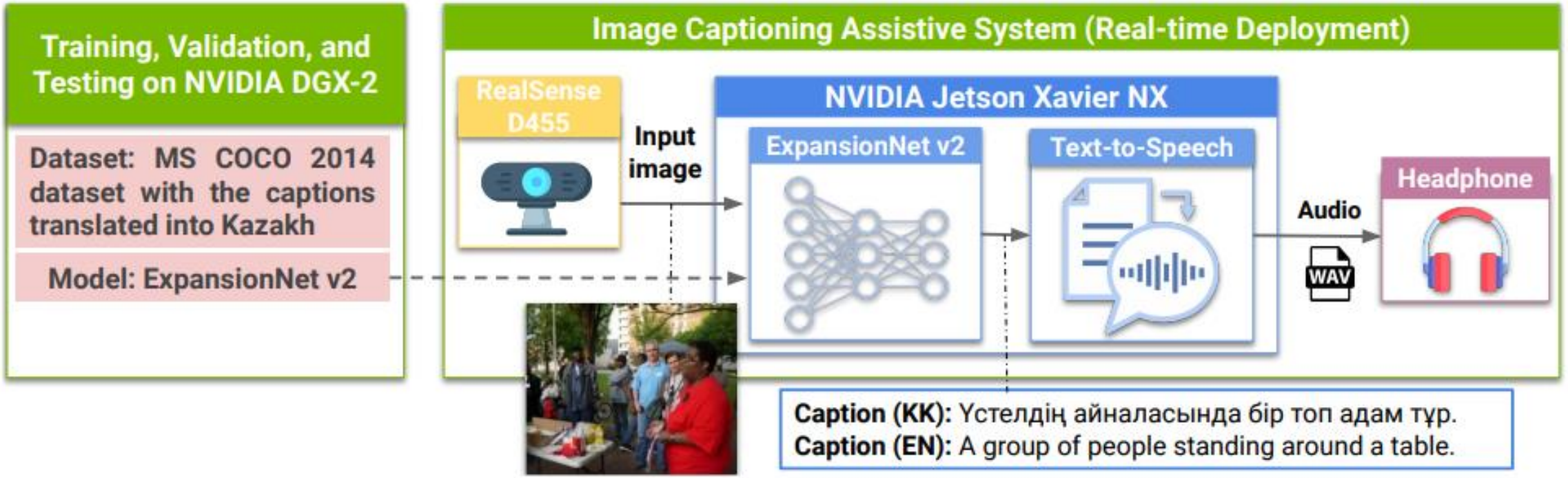
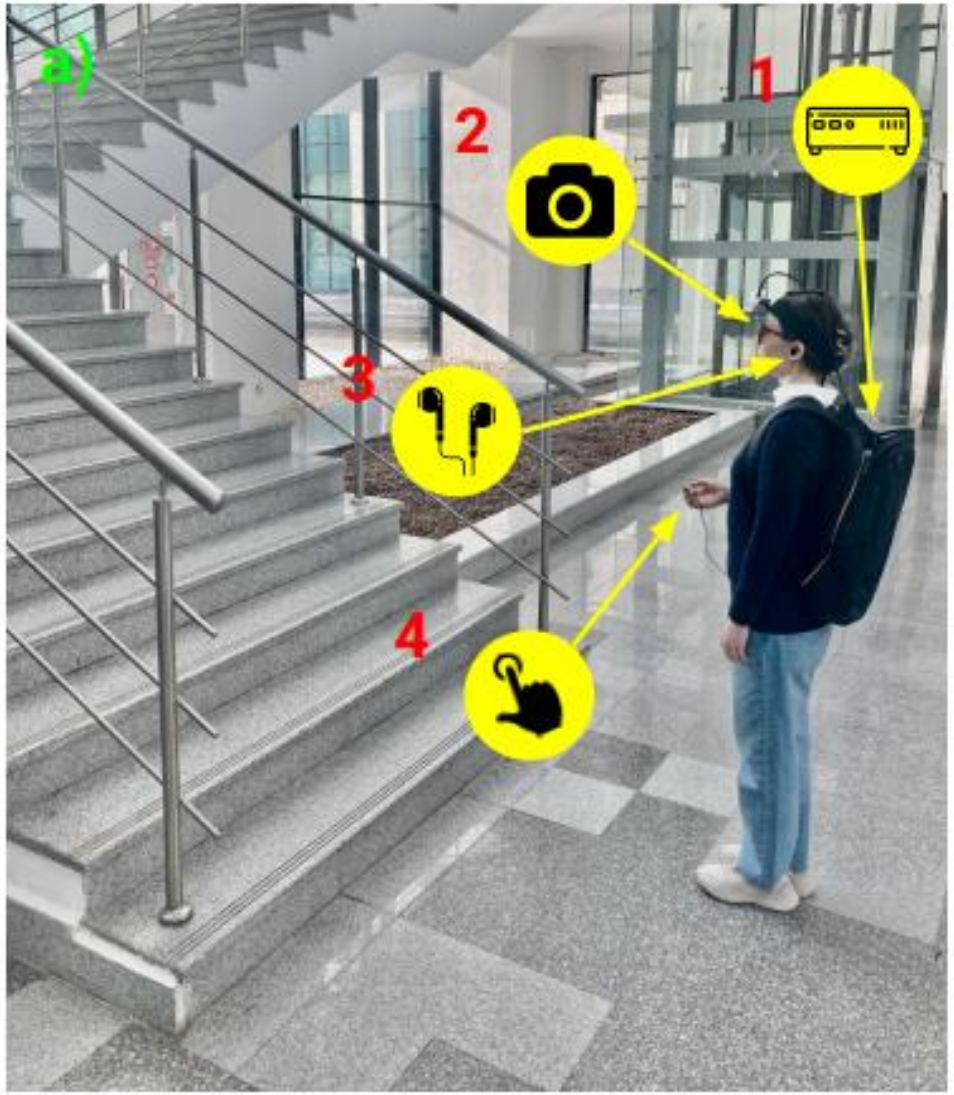


Image Captioning for the Visually Impaired and Blind



Kazakh: Баспалдақ пен баспалдақ бар баспалдақ.
English: A staircase with stairs and steps.

Kazakh: Үстінде қызыл белгі бар ақ есік.
English: A white door with a red sign on it.

Kazakh: Орындықтарда отырған екі әйел.
English: Two women sitting on chairs.

Kazakh: Еденде отырған сары белгі.
English: A yellow sign sitting on the floor.

Conclusions

- Advances in transformer-based language vision models enable natural-language interaction with AI.
- This opens new horizons to create assistive technologies for the cognitive and visual impairments.
- Training these models from scratch is often not feasible due to the training data size and computational requirements.
- Unfortunately, these models perform best for English language and their performance degrades for other languages [1].
- Therefore, efforts should focus on customization and fine-tuning of these models for Kazakh language.

[1] Seghier ML. ChatGPT: not all languages are equal. Nature. 2023 Mar;615(7951):216. doi: 10.1038/d41586-023-00680-3.

We can only move forward if we work together.

For research collaboration and other inquiries, please contact us at issai@nu.edu.kz

You can find papers related to the presented research at https://issai.nu.edu.kz/all-publications/