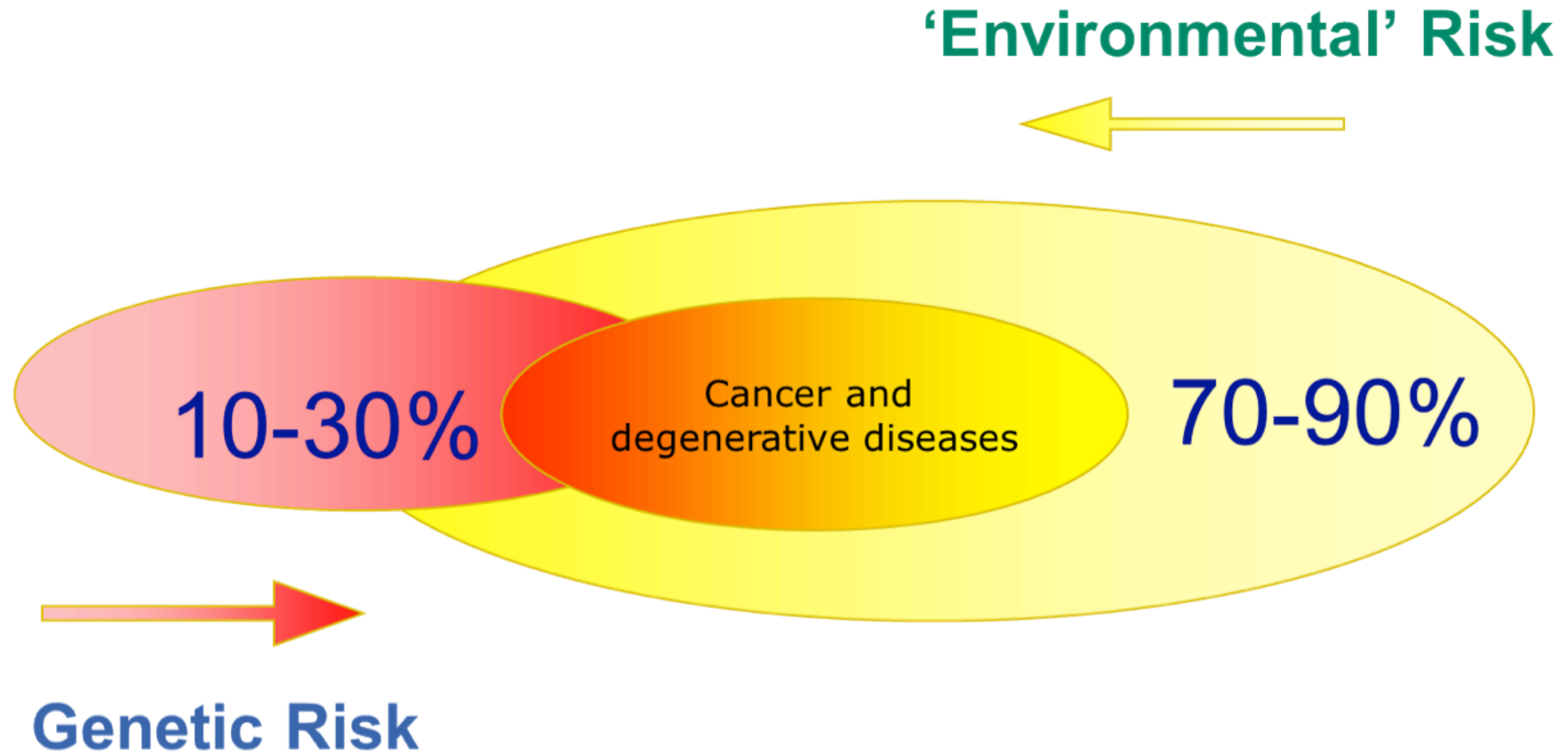
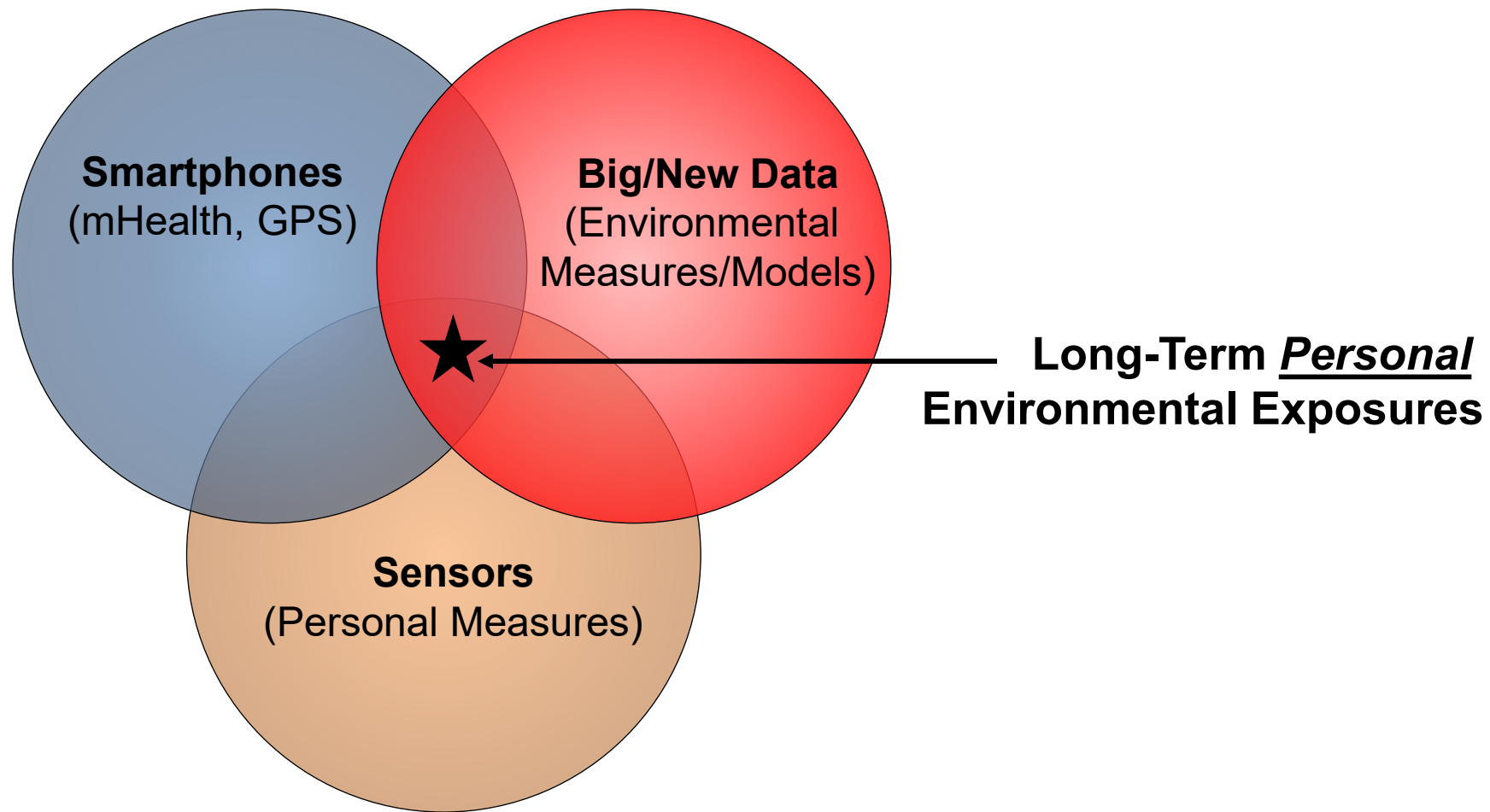


Revolutionizing Environmental Exposure Science with Artificial Intelligence

Perry Hystad
Professor, College of Health
Oregon State University

Why do we need better exposure measures?





Curr Envir Health Rpt (2017) 4:463–471
DOI 10.1007/s40572-017-0163-y



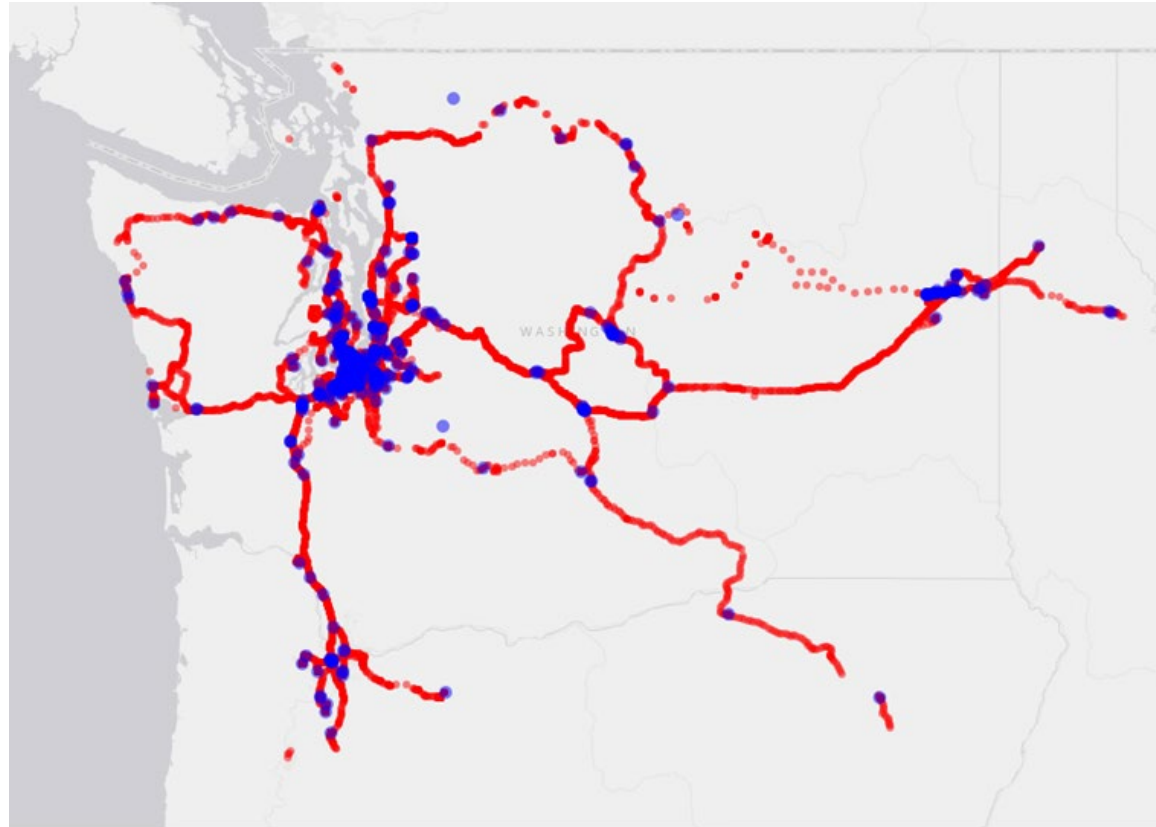
AIR POLLUTION AND HEALTH (S ADAR AND B HOFFMANN, SECTION EDITORS)

Towards Personal Exposures: How Technology Is Changing Air Pollution and Health Research

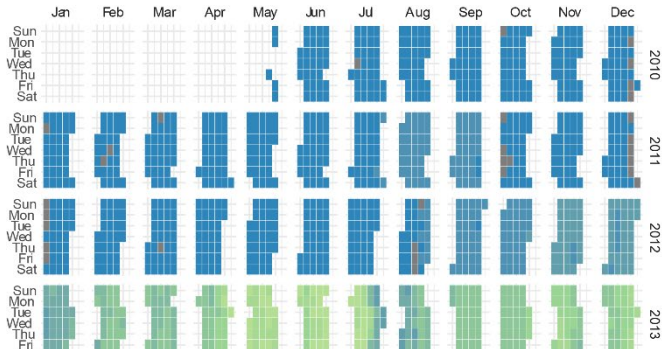
A. Larkin¹ • P. Hystad²

“Bring Your Own Location Data”: Use of Google Smartphone Location History Data for Environmental Health Research

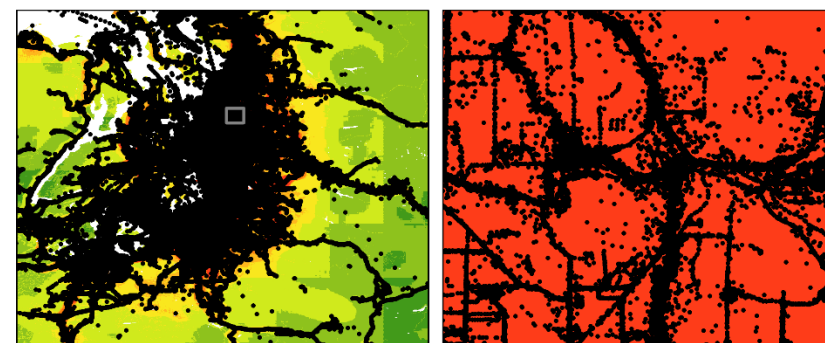
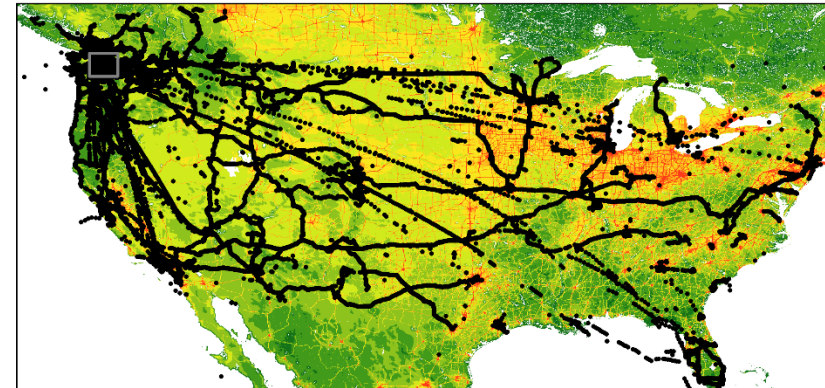
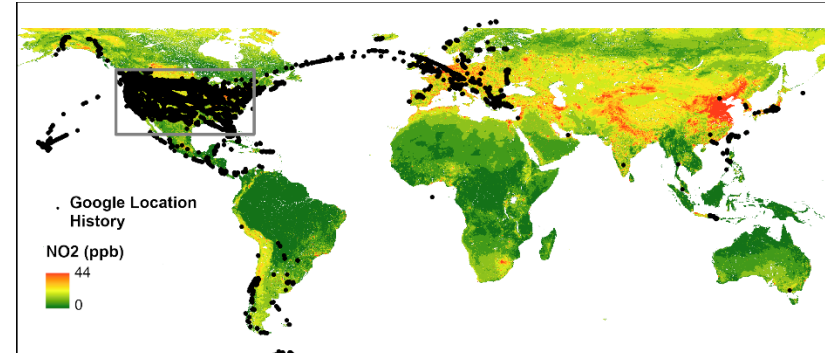
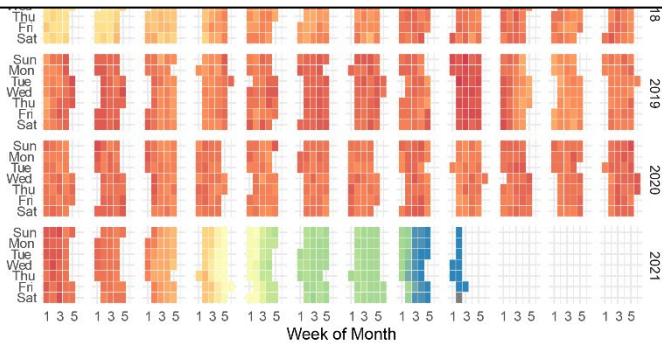
```
2021_MARCH.json - Notepad
File Edit Format View Help
{
  "timelineObjects": [ {
    "activitySegment": {
      "startLocation": {
        "latitudeE7": 445327917,
        "longitudeE7": -1232567430,
        "sourceInfo": {
          "deviceTag": -925283106
        }
      },
      "endLocation": {
        "latitudeE7": 445662992,
        "longitudeE7": -1232799340,
        "sourceInfo": {
          "deviceTag": -925283106
        }
      },
      "duration": {
        "startTimestampMs": "1614558022964",
        "endTimestampMs": "1614616867352"
      },
      "distance": 7989,
      "activityType": "IN_PASSENGER_VEHICLE",
      "confidence": "HIGH",
      "activities": [ {
        "activityType": "IN_PASSENGER_VEHICLE",
        "probability": 83.517324924469
      }, {
        "activityType": "STILL",
        "probability": 12.726157903671265
      }, {
        "activityType": "WALKING",
        "probability": 2.614324167370796
      }, {
        "activityType": "CYCLING",
        "probability": 0.4114963114261627
      }, {
        "activityType": "MOTORCYCLING",
        "probability": 0.3070661099627614
      }, {
        "activityType": "IN_BUS",
        "probability": 0.1503273961134255
      }, {
        "activityType": "IN_FERRY",
        "probability": 0.09590949630364776
      }, {
        "activityType": "IN_TAXI",
        "probability": 0.0000000000000000
      }
    ]
  }
]
```

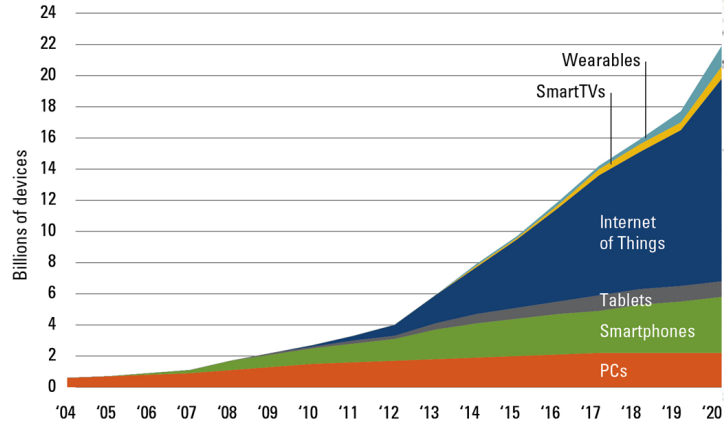
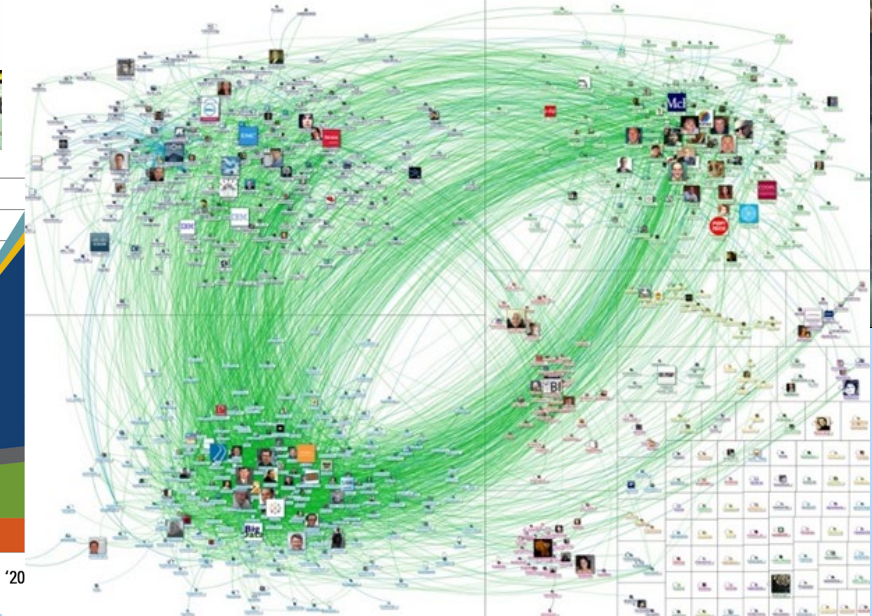
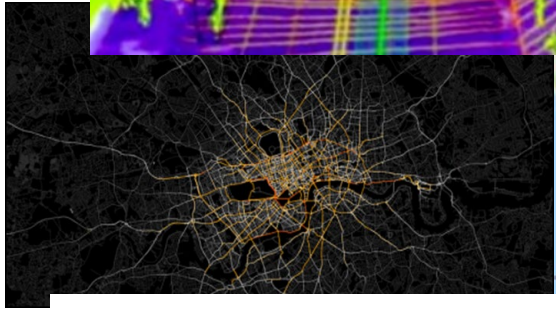
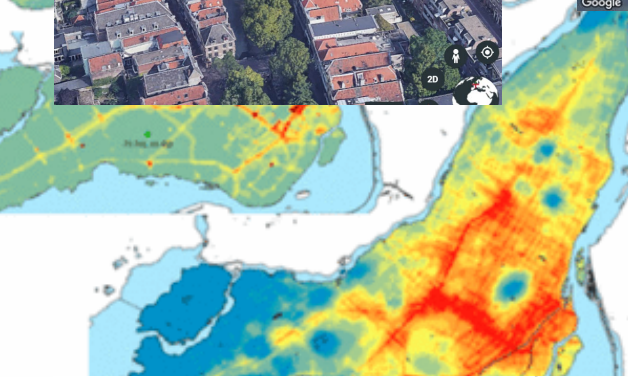
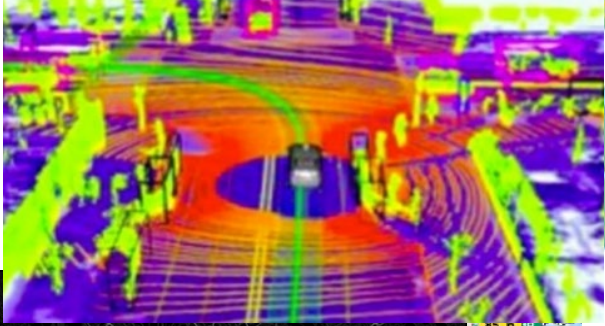
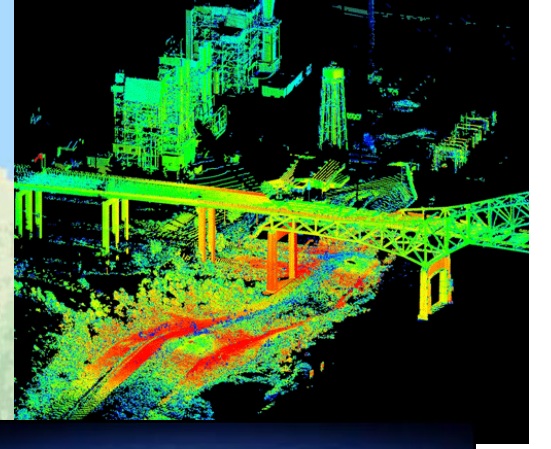
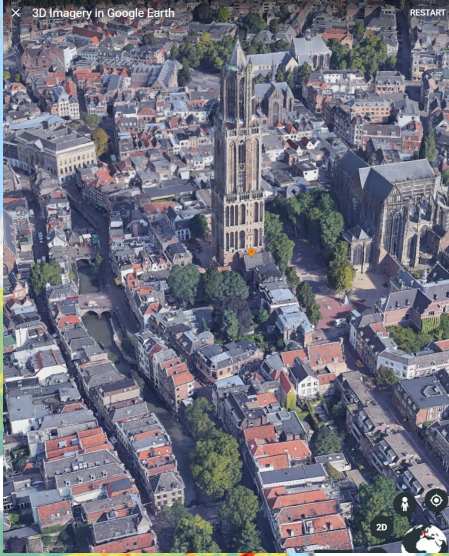


“Bring Your Own Location Data”: Use of Google Smartphone Location History Data for Environmental Health Research



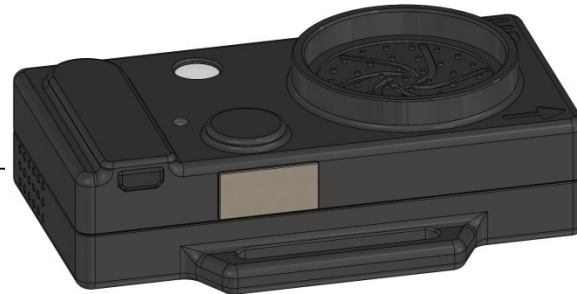
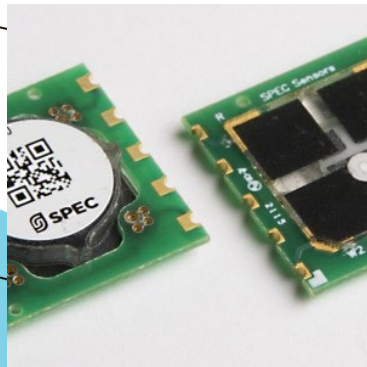
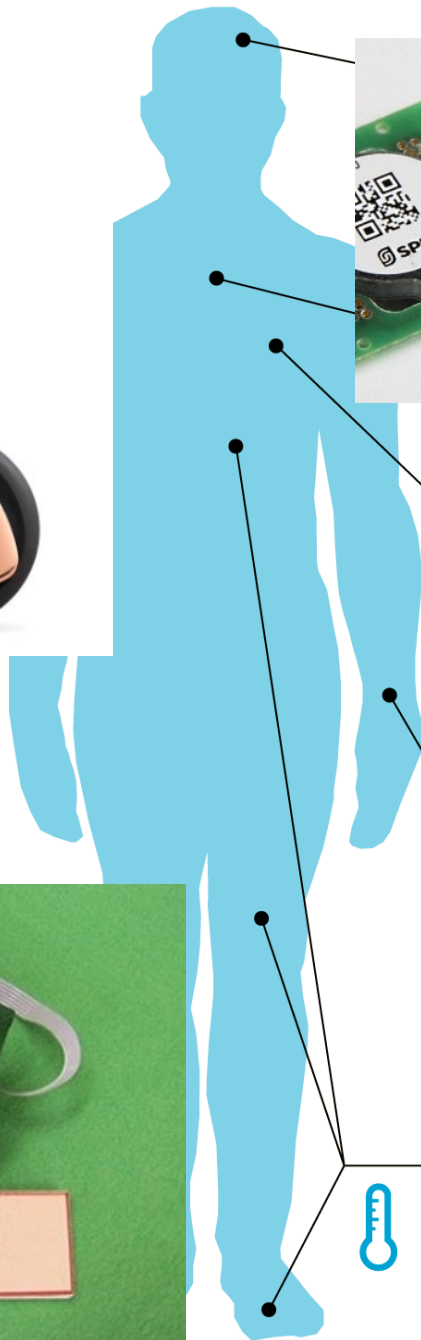
- 400 Twins provided data
- Average 58 months! of time-activity data.
- 370 million time-location points.





Sources: Gartner, IDC, Strategy Analytics, Machina research, company filings, BII estimates





Accelerometer

Altimeter

Digital camera



Electrocardiogram

Electromyography

Electroencephalogram

Electrodermal activity



Smartwatches



Sensors



Location GPS



Microphone



Oxygen



Bluetooth



Pressure



Thermometer

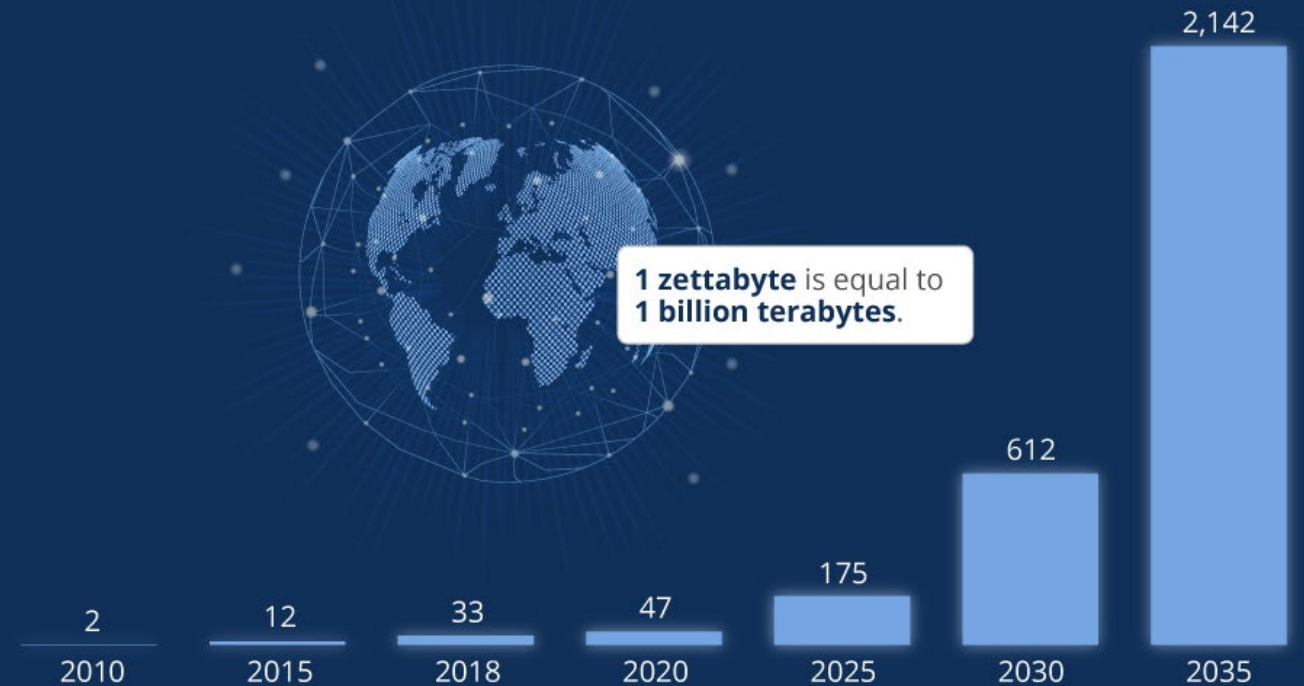




Data Overload!

Global Data Creation is About to Explode

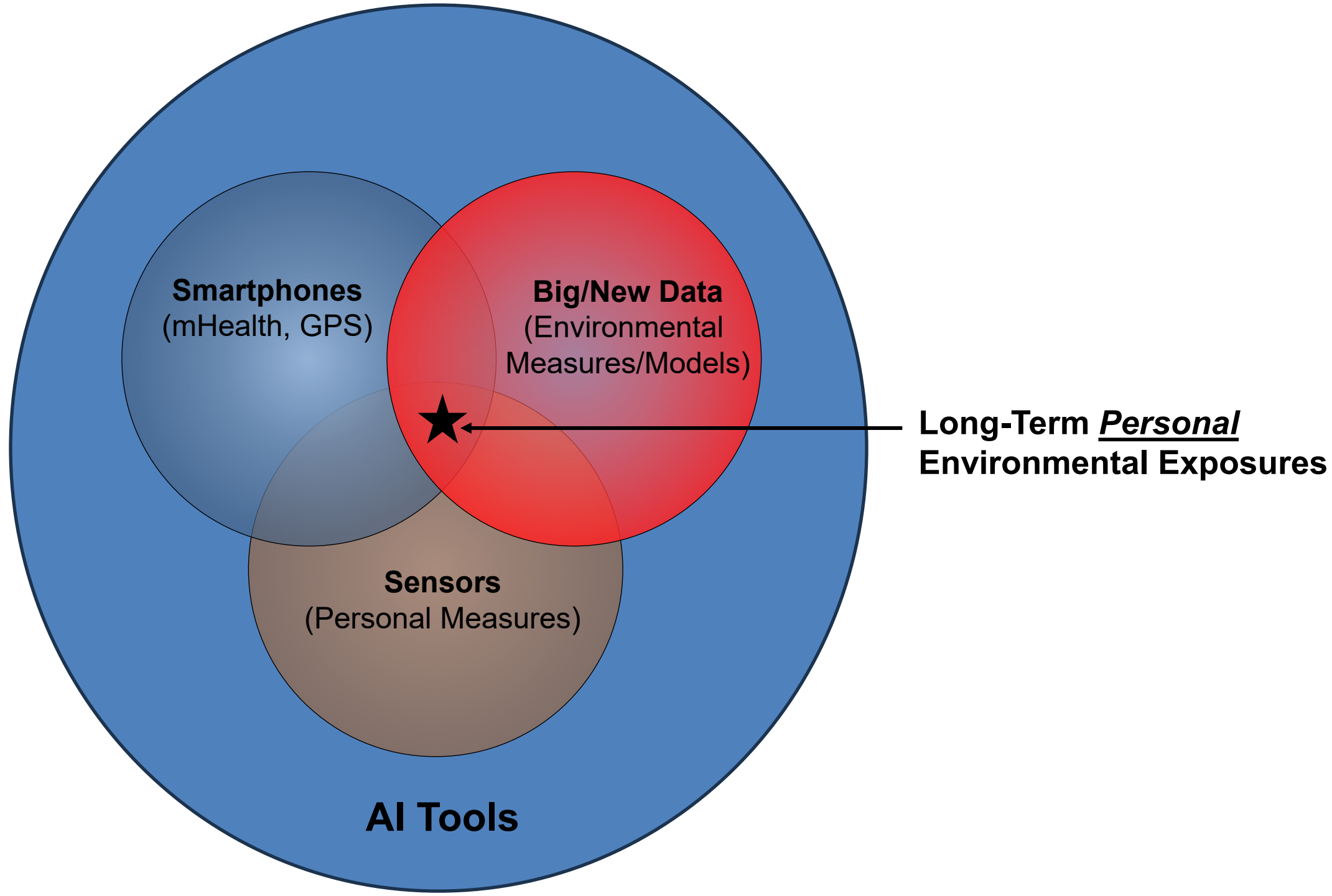
Actual and forecast amount of data created worldwide 2010-2035 (in zettabytes)



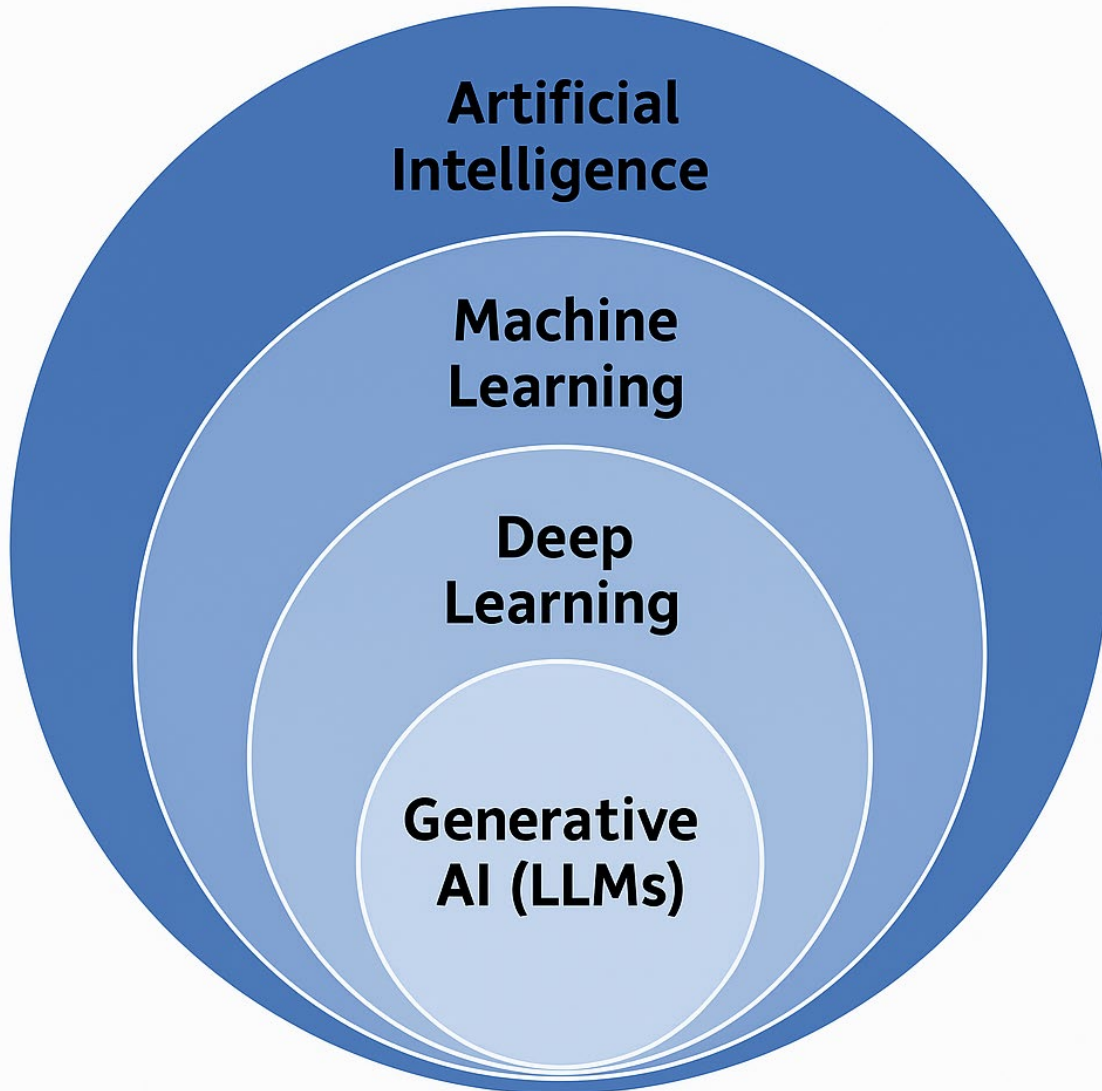
@StatistaCharts

Source: Statista Digital Economy Compass 2019

statista



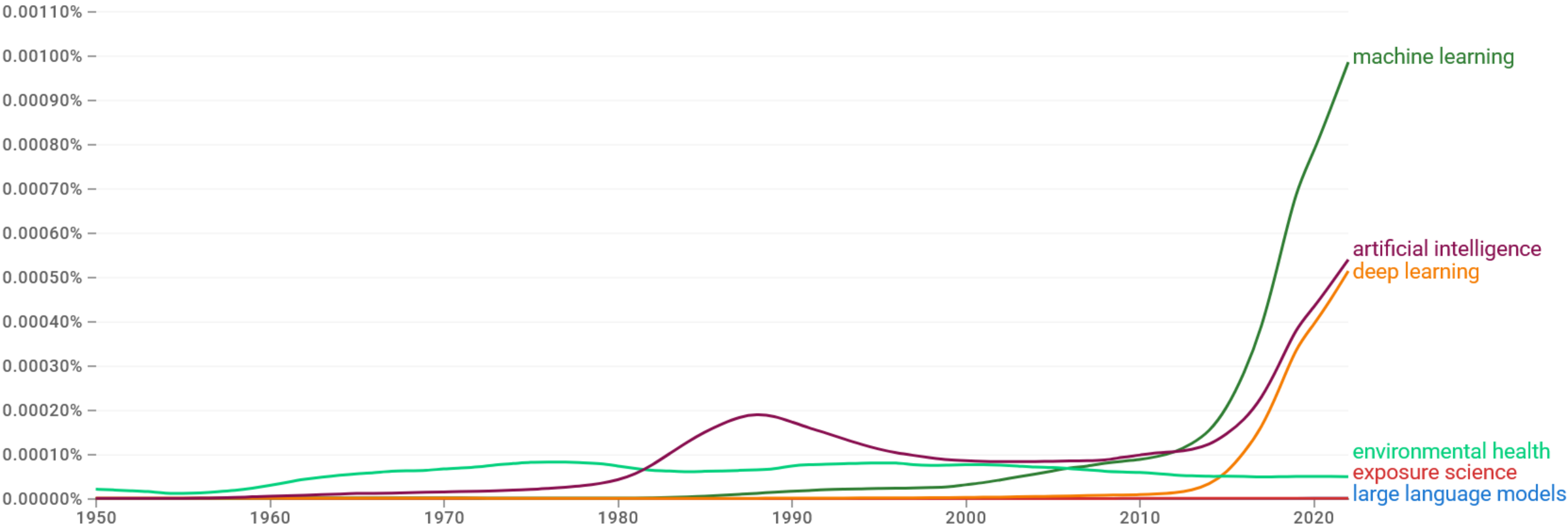
AI Toolkit for Exposure Science



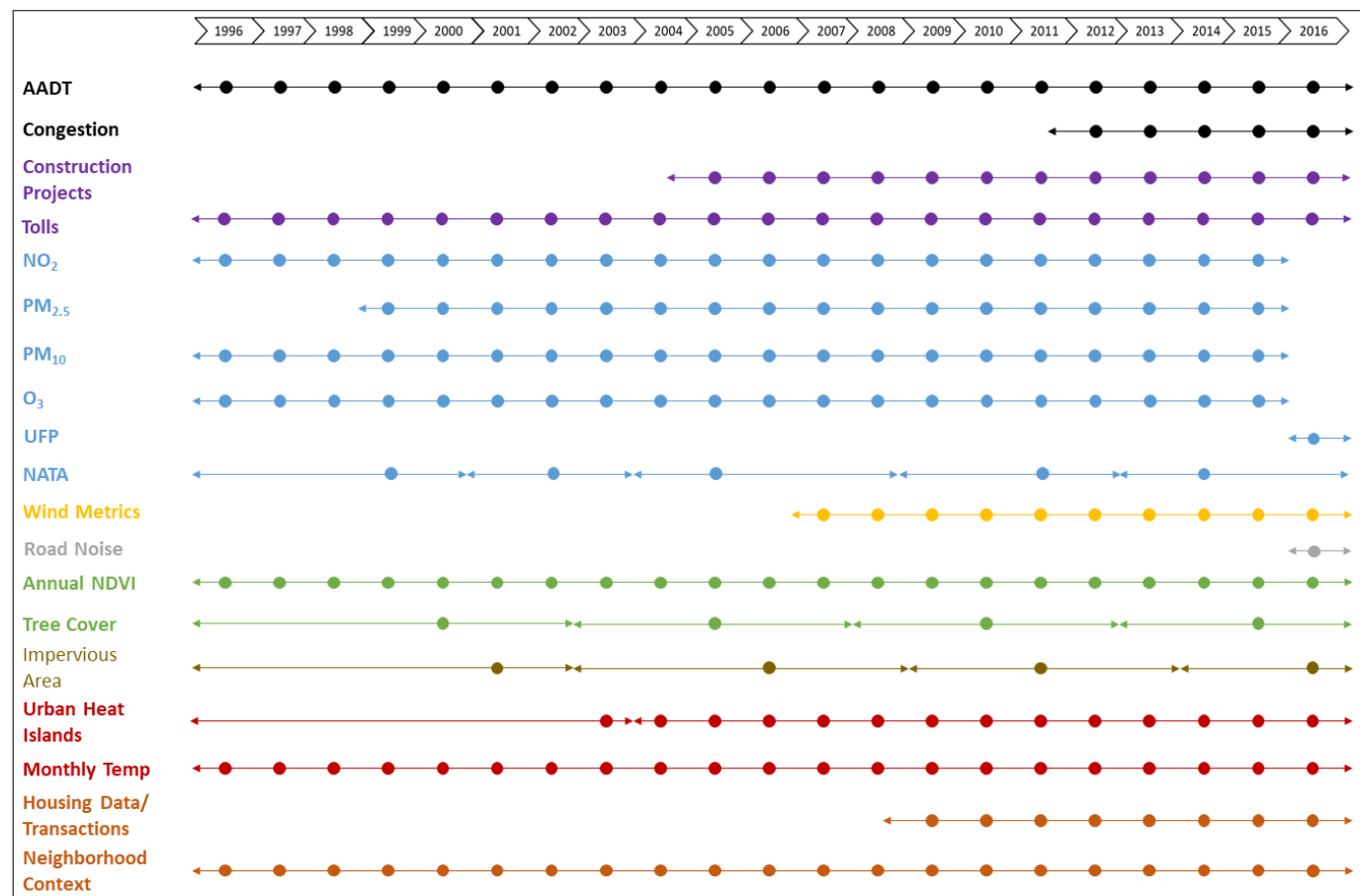
- **AI:** The science of creating systems that can perform tasks requiring human intelligence, such as reasoning, learning, or decision-making.
- **Machine Learning:** A subset of AI where systems learn patterns from data and improve their performance without being explicitly programmed.
- **Deep Learning:** A specialized form of machine learning that uses multi-layered neural networks to model complex patterns in large datasets.
- **Generative AI:** A type of deep learning model that generates new content, such as text, images, or audio, based on learned patterns from training data.

large language models,exposure science,machine learning,deep learning,environn

1950 - 2022 English Case-Insensitive Smoothing



AI is not just analyzing exposure data – it's redefining what is available for exposure data



Built E
Infra



cod
ent

HEI
Health Effects Institute

NUMBER 223
FEBRUARY 2025

RESEARCH REPORT

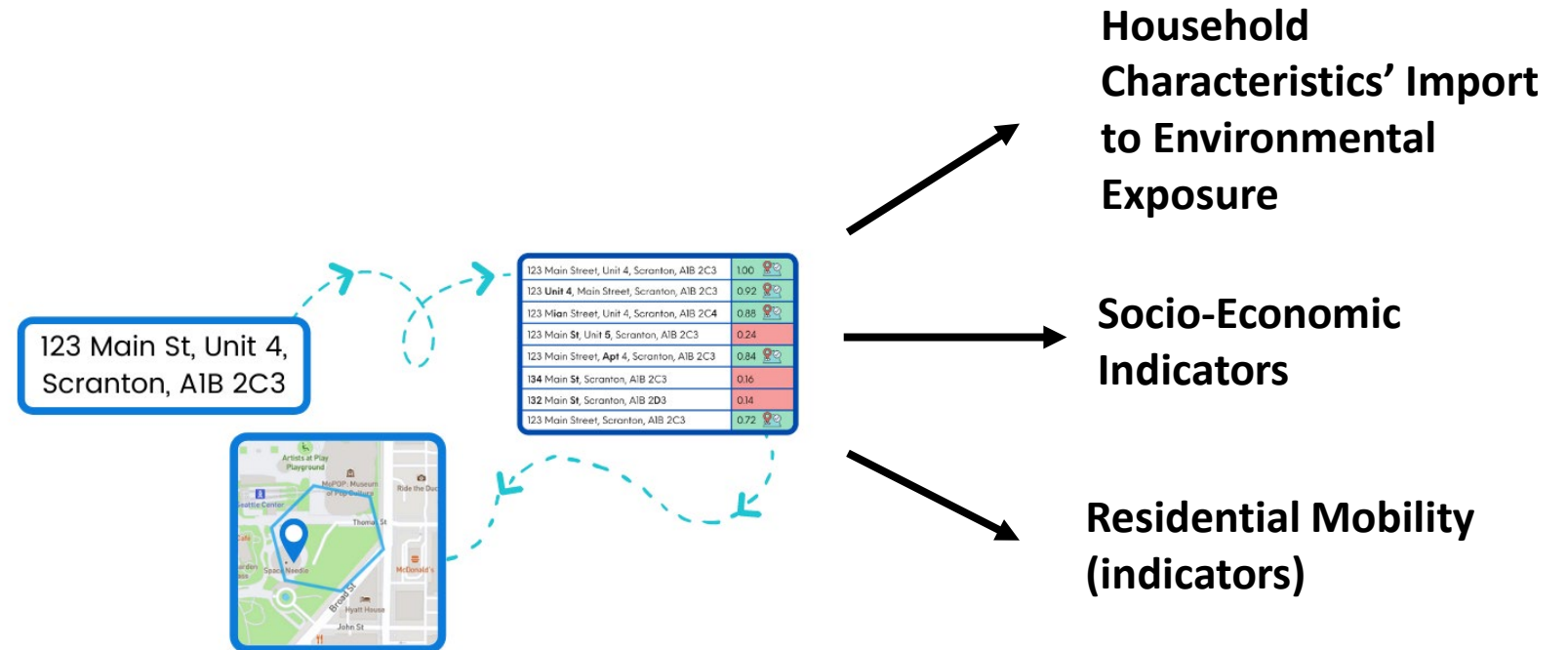
Impacts of Vehicle Emission Regulations and Local Congestion Policies on Birth Outcomes Associated with Traffic Air Pollution

Perry Hystad, Mary Willis, Elaine Hill, David Schrank, John Molitor, Andrew Larkin, and Beate Ritz

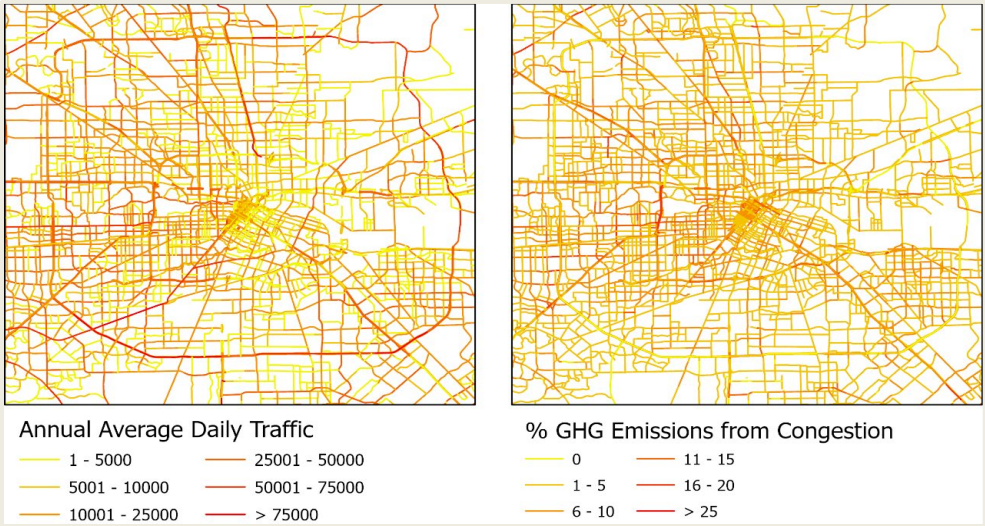
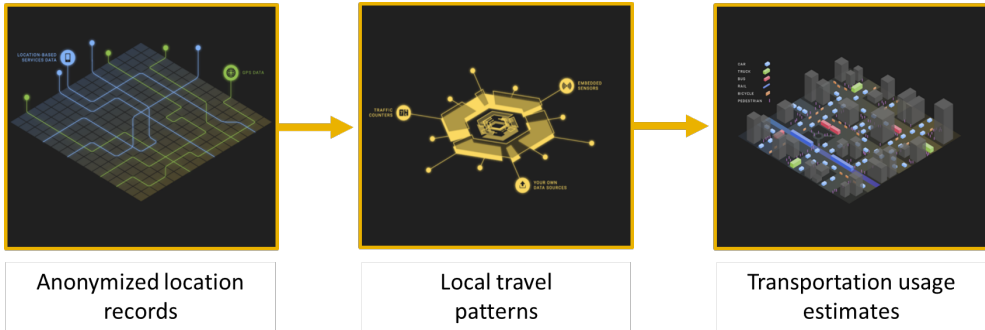
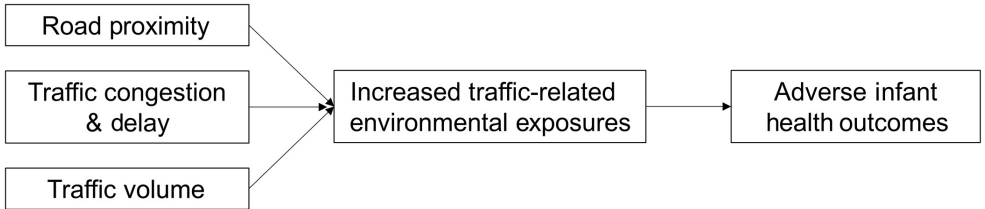
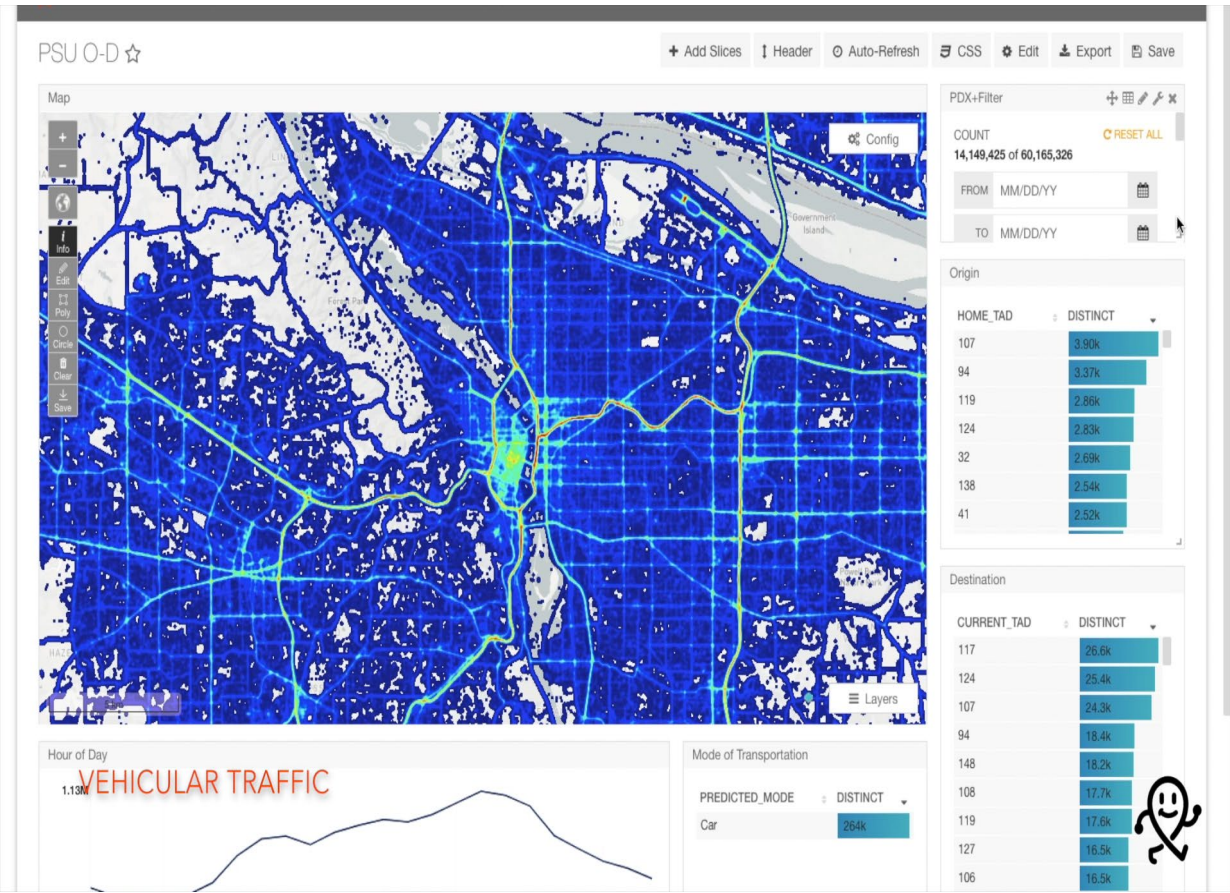
INCLUDES A COMMENTARY BY THE INSTITUTE'S REVIEW COMMITTEE

www.healtheffects.org

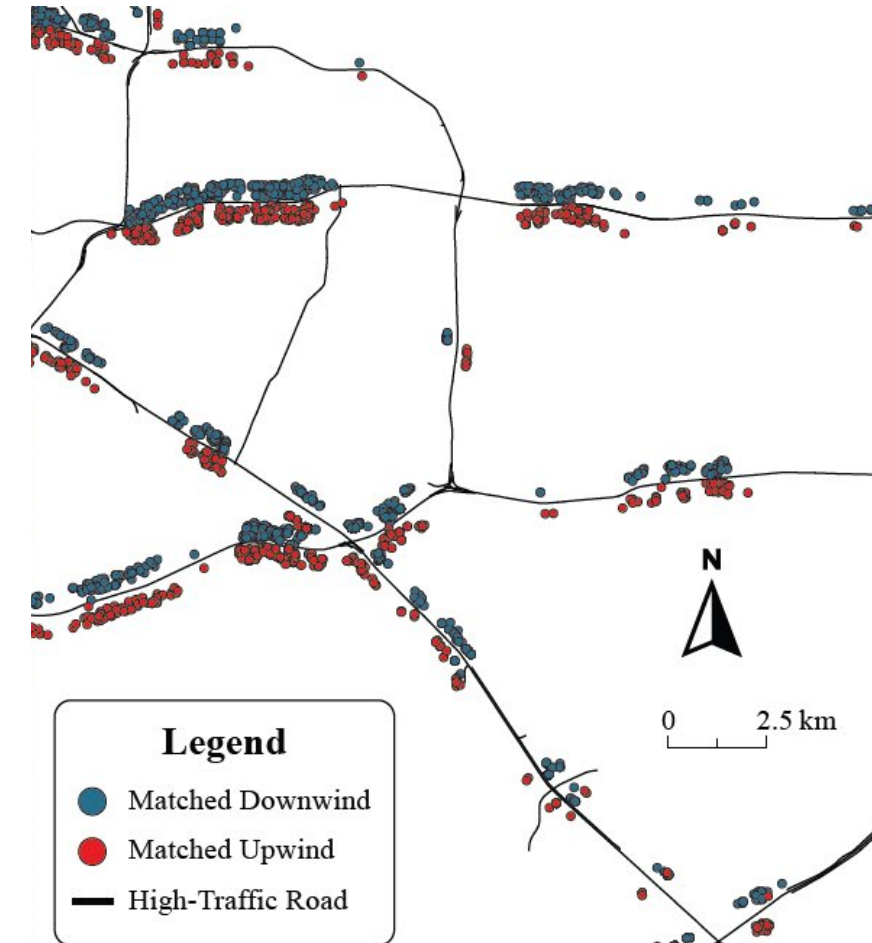
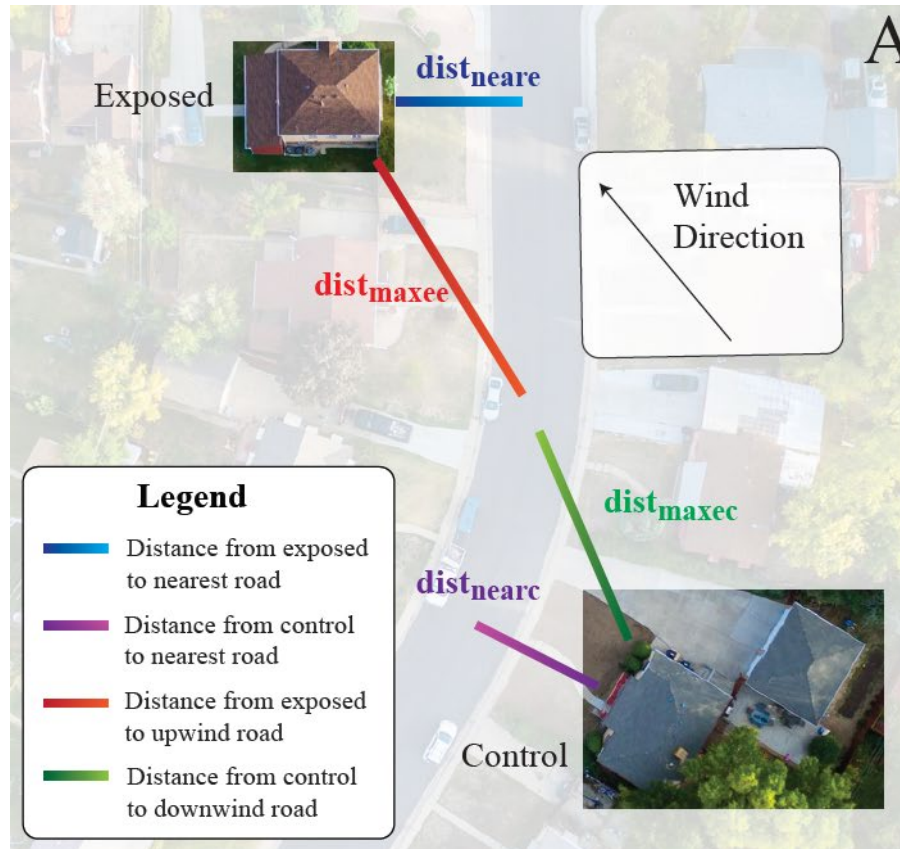
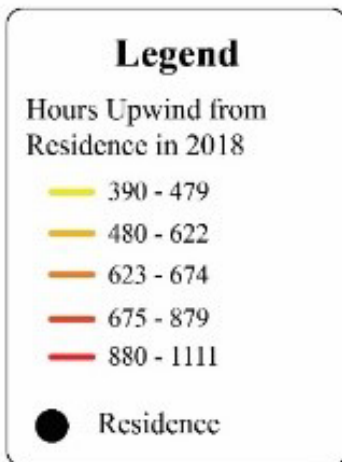
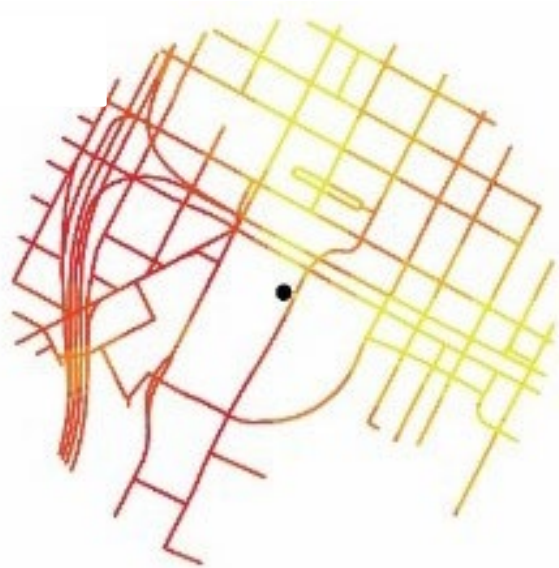
Integrating Housing Characteristics



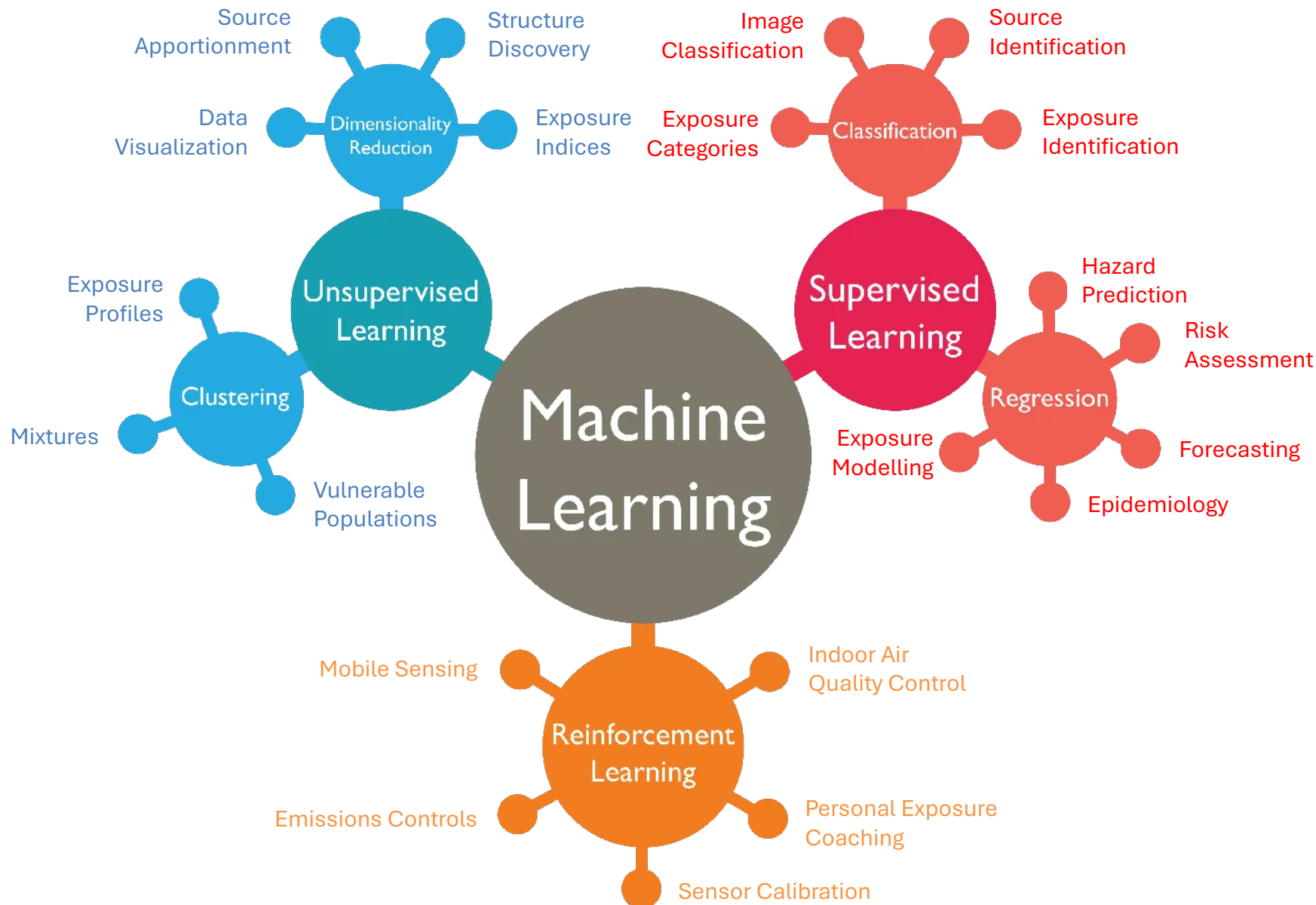
Measuring Congestion-Related Air Pollution



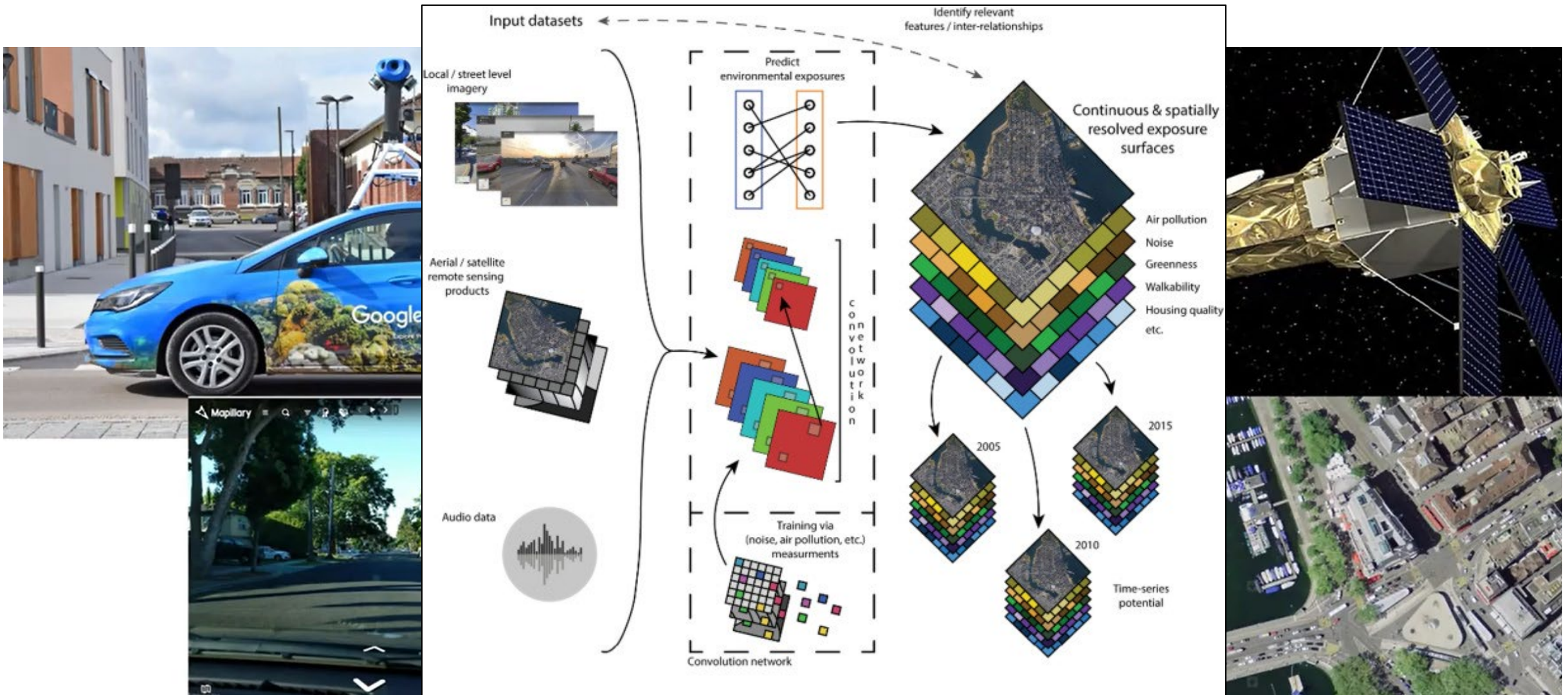
Matching Individuals Upwind and Downwind of the Same High-Traffic Road



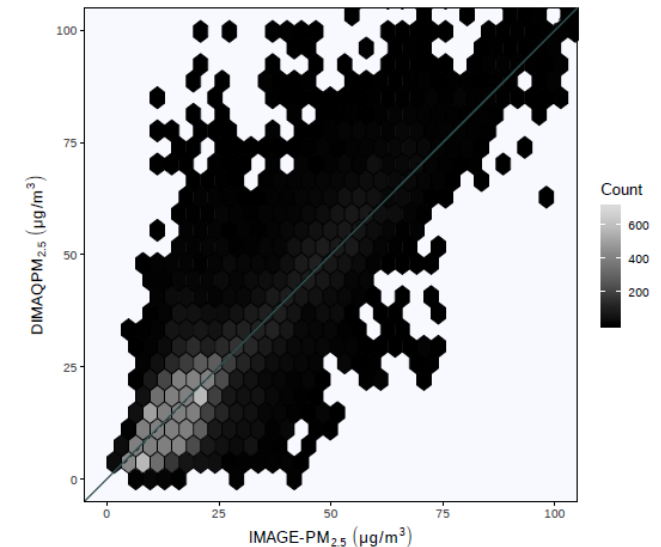
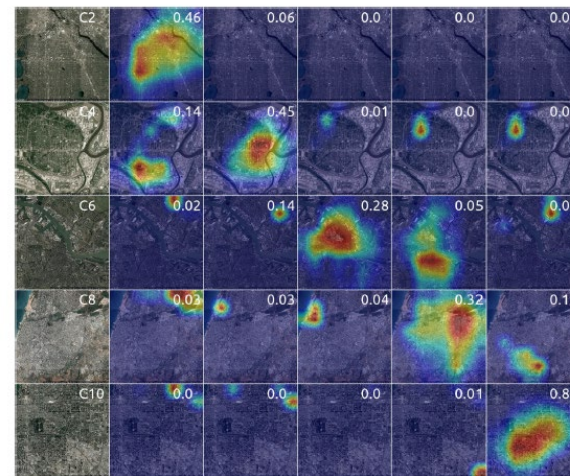
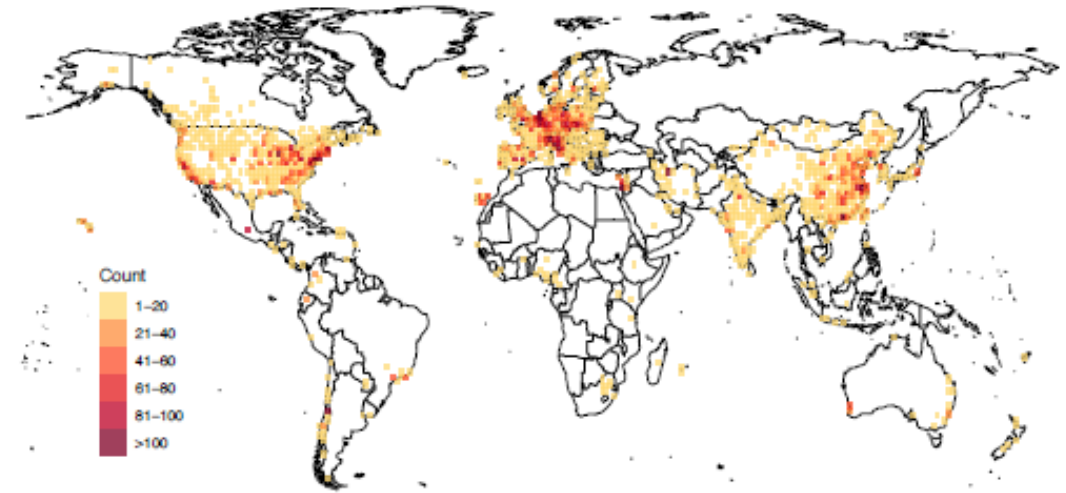
ML Central to Environmental Exposure Science



Explosion of Images Available for Exposure Science



Satellite/Aerial Imagery Exposure Models



Street-View Exposure Models




Predicting Complex Exposure Measures


2. Which street has a higher quality of nature?

Move the slider left or right to make your choice. Move the slider farther if you feel more strongly about your choice. You cannot rank the images as equal.

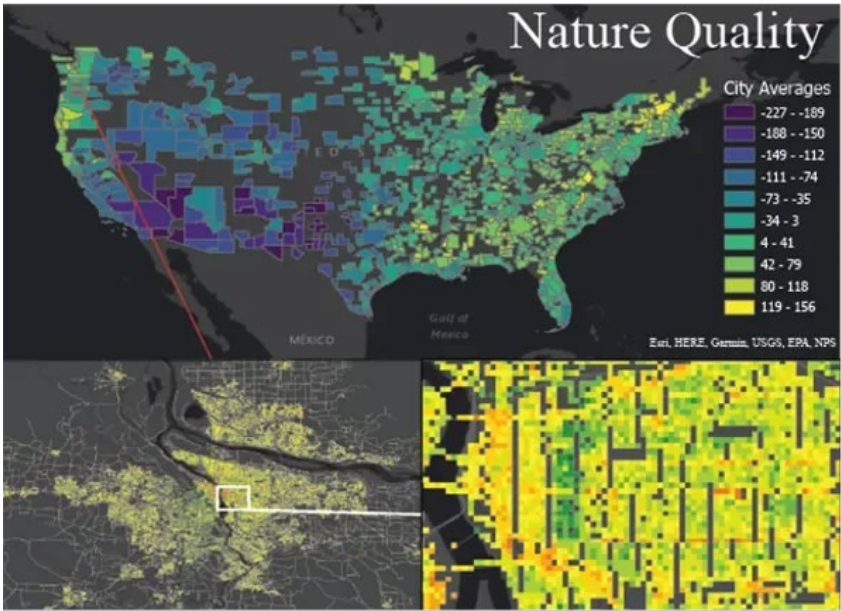
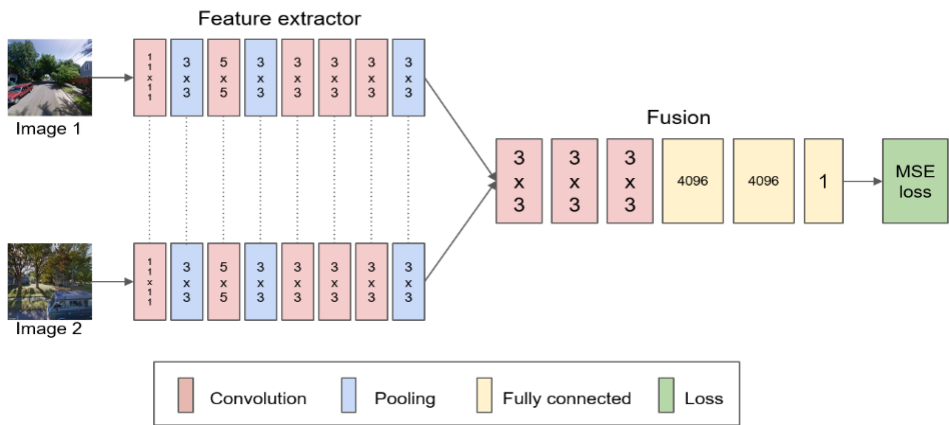
left street has a higher quality of nature

right street has a higher quality of nature





Same composition, different quality



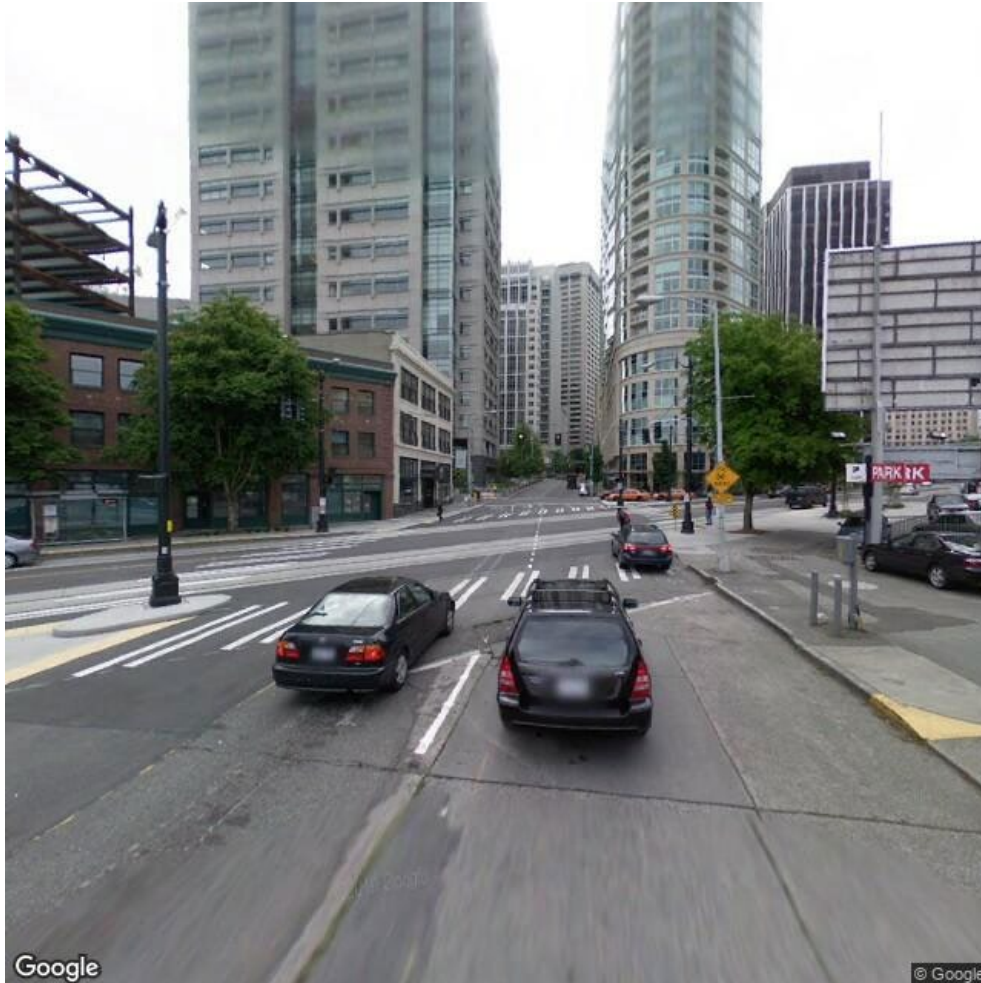
City- and Census Tract-Level Scores



- A deep learning model trained on vast text data.
- ChatGPT, Gemini, Claude, LLaMA, DeepSeek, etc.
- LLMs extend what ML/DL can do by offering multimodal integration at scale.
- Capable of integrating, reasoning, summarizing, answering questions, and making predictions.

LLMs Moving Extremely Fast!

How would you rank the quality of nature in this street? Rank from 0 (lowest quality) to 100 (highest quality). Explain why you chose the specific score.



I would rank the **nature quality** of this street as **25 out of 100**.

Explanation:

Positive indicators:

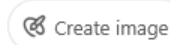
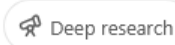
- There are **some mature street trees** visible along both sides of the street, which provide minimal shade and a small visual buffer from the built environment.
- The trees appear relatively healthy and integrated into the sidewalk environment.

Negative indicators:

- The **built environment dominates** the scene—tall buildings, wide roadways, and concrete surfaces occupy the majority of visual and spatial space.
- There is **no visible green space**, parks, or landscaping beyond the street trees.
- The **street design is car-centric**, with multiple lanes and little evidence of natural elements incorporated into pedestrian or public space design.
- **Lack of biodiversity**—only one or two tree species are visible, with no shrubs, flowers, or grass.
- **No visible water features, soil, or ecological elements** that would support habitat or environmental function.



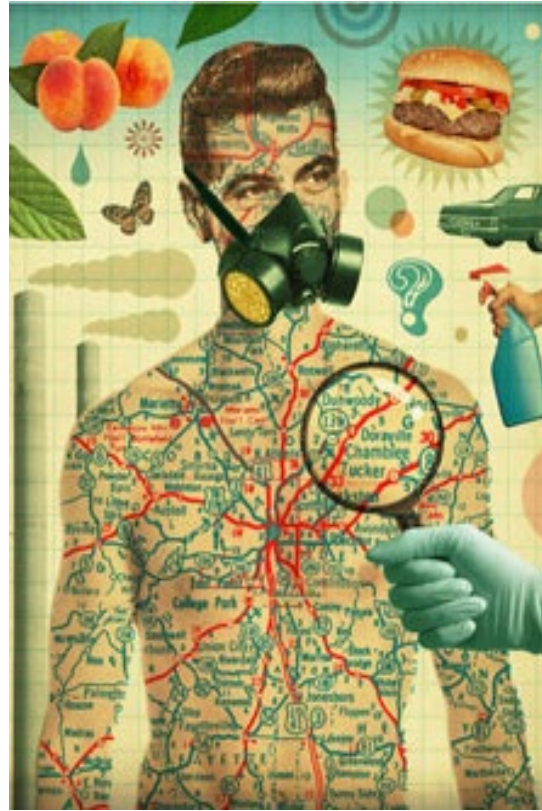
Ask anything



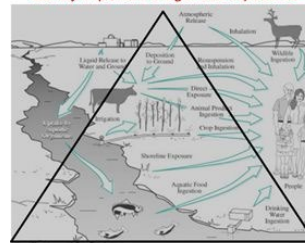
LLMs as a Tool to Operationalize the Exposome

External Exposome

- Contextual factors
- SES
- Climate
- Physical activity
- Diet
- Air pollution
- Occupational exposures
- Noise
- Radiation
- Chemical exposures
- Green space
- Etc.

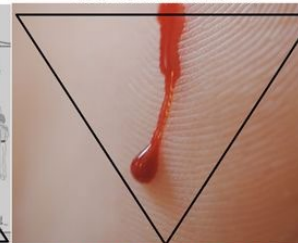


Bottom-up Exposomics
Identify important exogenous exposures



Measure chemicals in air, water & food

Top-down Exposomics
Measure chemicals in blood



Identify all important exposures

Internal Exposome

- Inflammation
- Oxidative stress
- Metabolomics
- Transcriptomic
- Proteomics
- Epigenetics
- Microbiome
- Lip peroxidation
- Immunomics
- Adductomics
- Etc.

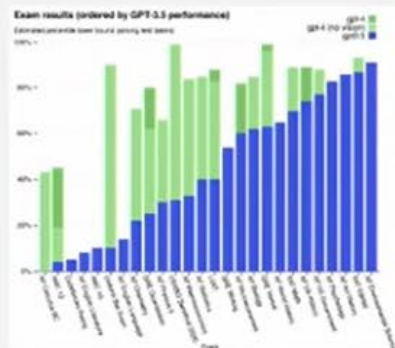
Foundation Models in Different Domains

Natural Language Processing

Stanford
Alpaca



Stanford Alpaca



ChatGPT/GPT-4 (OpenAI, 2023)

Computer Vision



Imagen (Saharia et al. 2022)



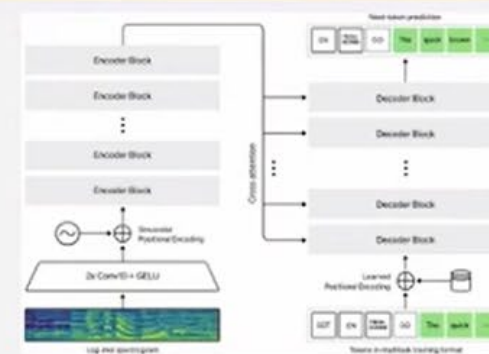
Segment Anything (Kirillov et al, 2023)

Reinforcement Learning



Gato (Reed et al. 2022)

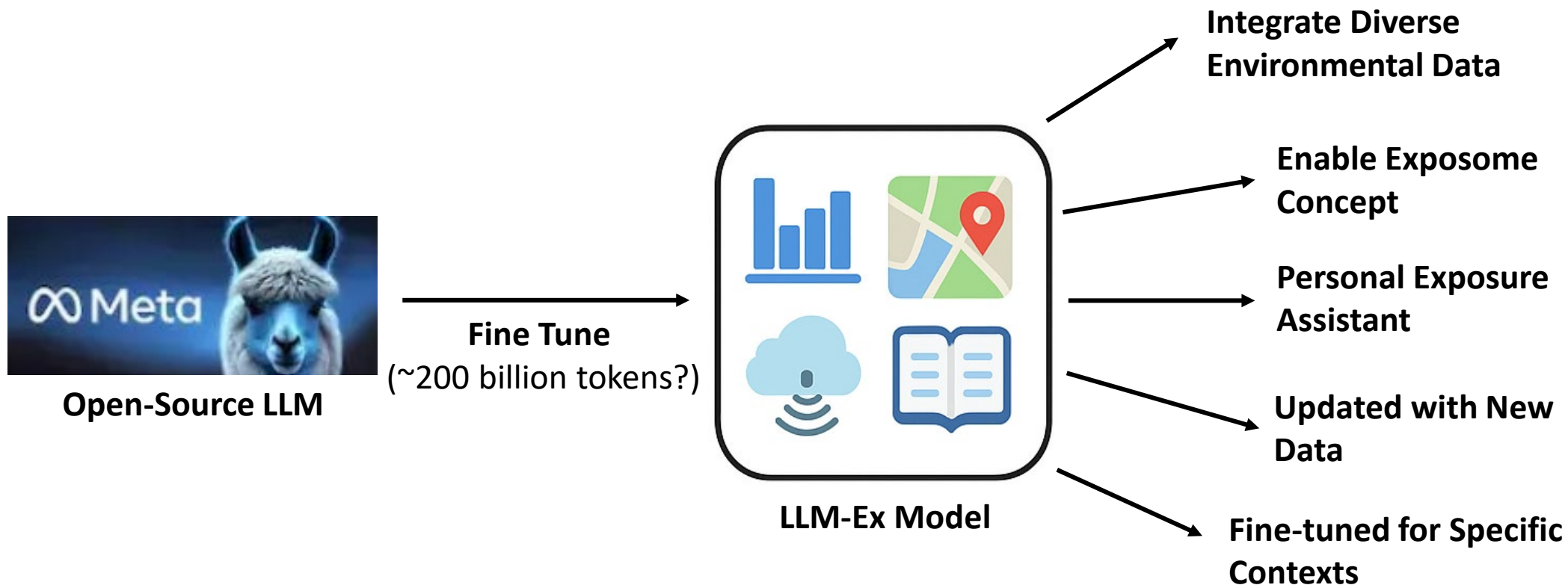
Signal Processing



Whisper (Radford et al. 2022)

Large-scale models designed for flexible adaptation across a wide range of downstream tasks.

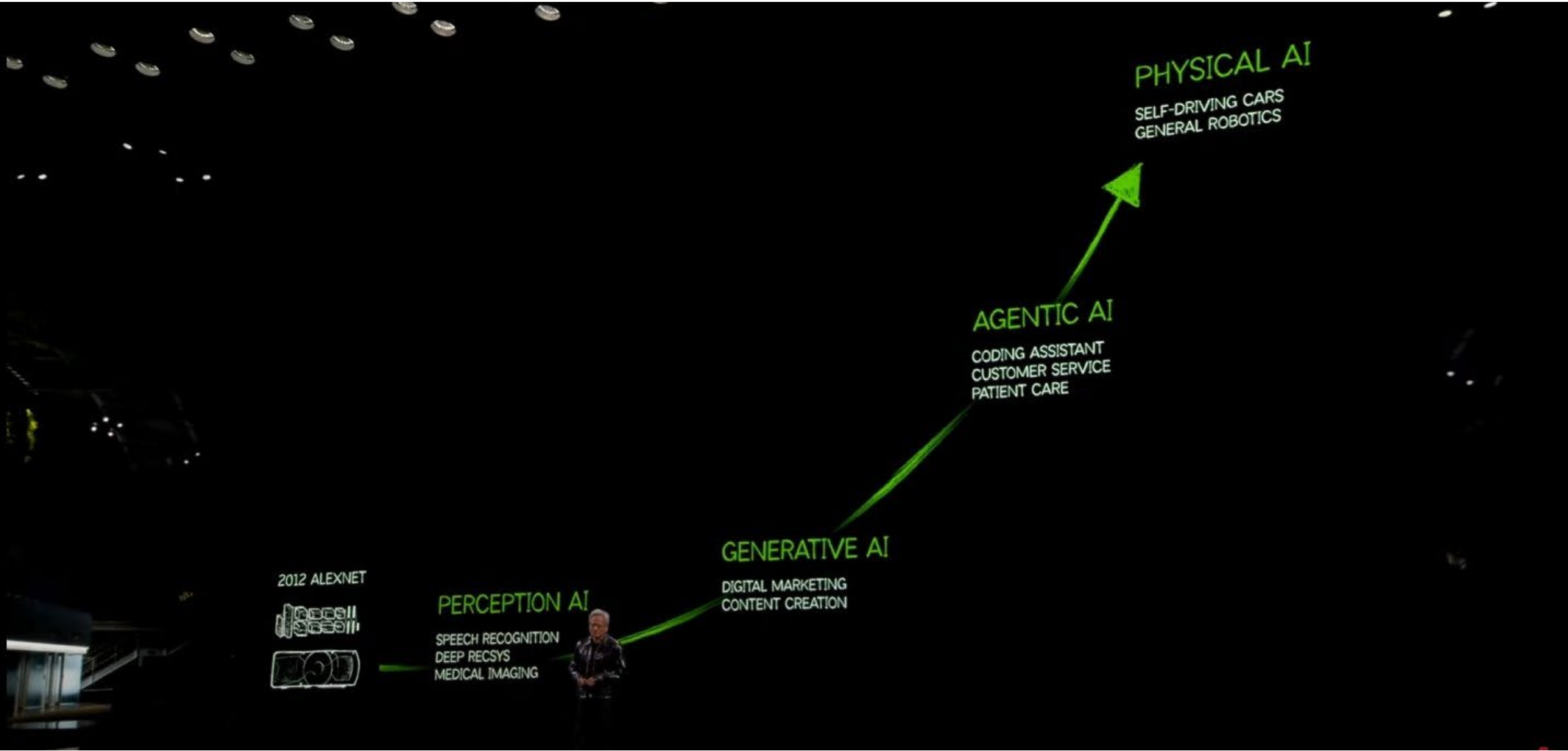
Foundational Model for Environmental Exposure Science: LLM-Ex



Ethical and Responsible Use Considerations

1. Direct environmental impact of LLM compute.
2. Copyright of data used to train models.
3. Privacy and confidentiality.
4. Bias and representativeness of environmental data.
5. Explainability of model predictions.
6. Misinformation (hallucinations).
7. Attribution of ideas.
8. Overdependence on automation.

Collaboration between computer scientists and environmental health experts is vital for optimal AI





Thank you – Questions?

Perry Hystad, perry.hystad@oregonstate.edu