

Alignment & Robotics (1)

AI 109

Last Time

- AI Alignment

Today

- Review of ***Alignment***
 - Getting AIs to agree with us about moral questions.
 - Making sure AI systems don't do things we consider morally wrong.
- Starting Robotics

InstructGPT (2022)

Step 1

Collect demonstration data, and train a supervised policy.

A prompt is sampled from our prompt dataset.

🧠 Explain the moon landing to a 6 year old

A labeler demonstrates the desired output behavior.

👤
✍️ Some people went to the moon...

This data is used to fine-tune GPT-3 with supervised learning.

🧠 SFT
✍️ 📄 📄 📄

Step 2

Collect comparison data, and train a reward model.

A prompt and several model outputs are sampled.

🧠 Explain the moon landing to a 6 year old

Ⓐ Explain gravity... Ⓑ Explain war...
Ⓒ Moon is natural satellite of... Ⓓ People went to the moon...

A labeler ranks the outputs from best to worst.

👤
Ⓓ > Ⓒ > Ⓐ = Ⓑ

This data is used to train our reward model.

🧠 RM
Ⓓ > Ⓒ > Ⓐ = Ⓑ

Step 3

Optimize a policy against the reward model using reinforcement learning.

A new prompt is sampled from the dataset.

🐸 Write a story about frogs

The policy generates an output.

🧠 PPO
Once upon a time...

The reward model calculates a reward for the output.

🧠 RM

The reward is used to update the policy using PPO.

r_k

Step 1: Supervised Fine-tuning

- Also called **instruction fine-tuning**
- Start from the base model.
- Do supervised training using **demonstrations** of tasks.
- This is a kind of **imitation learning**.
- Examples:
 - <https://huggingface.co/datasets/nvidia/Nemotron-Cascade-2-SFT-Data/viewer/chat/train?row=0>

Step 2: Reward Modeling

- First, collect a data set of **comparisons** between two responses x and y for the same prompt
- Then one of two methods
 - Old: have humans rank responses.
 - New: have an AI rank responses.
- Train a neural network $r(x,y)$ that gives preferred answers a higher rating.
 - This is called the **reward model**.

Step 3: RLHF

- First: sample a set of prompts from a data set.
- Second: have your model generate completions.
- Third: score each completion with your reward model
- Fourth: update the model using reinforcement learning on the scores from the reward model.

Why does alignment matter?

Alignment Today

- <https://www.anthropic.com/glasswing>

Robotics

What is a Robot?

- Surprisingly hard to define.
- Let's look at several things that might be robots.

Robots



industrial robots

Robots



P3-DX

roomba



small wheeled robots

Robots

3DRobotics Solo



small drones

Robots

MQ-9 Reaper



large drones

Robots

Stanley (at the Smithsonian down the road)



autonomous cars

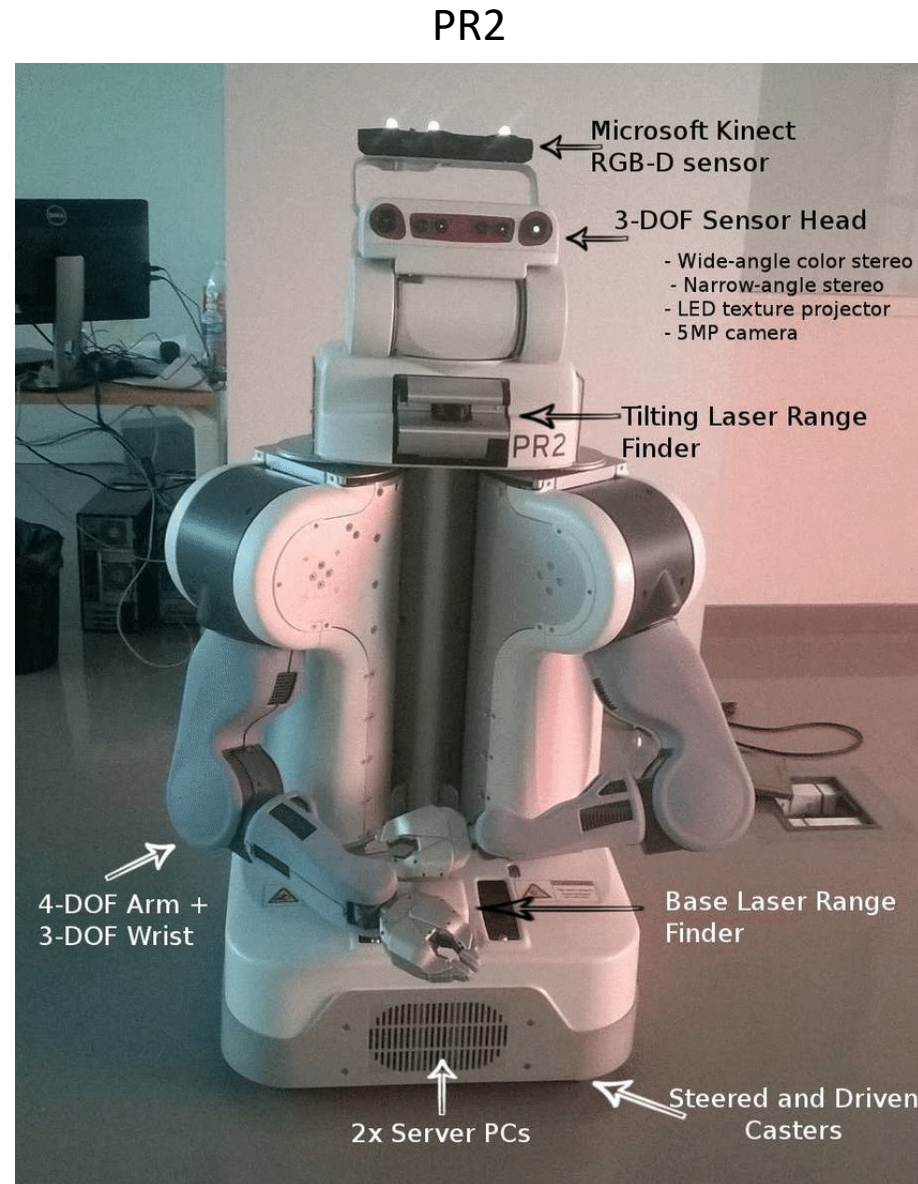
Robots

Nao



small humanoids

Robots



wheeled humanoids

Robots

Tesla Optimus



humanoids

Robots

Hiroshi Ishiguro's robots



androids

Robot?



What do these robots have in common?

Parts of Robotics

- Perception
 - How to make a robot see.
- Planning
 - How to make a plan to move through space.
- Control
 - How to make the robot's motors

Control

- Robots are made of motors.
 - Convert electrical energy to mechanical energy (spinning)
- **Control** is about determining how motors should spin.
 - Usually based on some kind of sensor input.
- Oldest part of robotics.
 - Dates to the 1800s with steam engines.



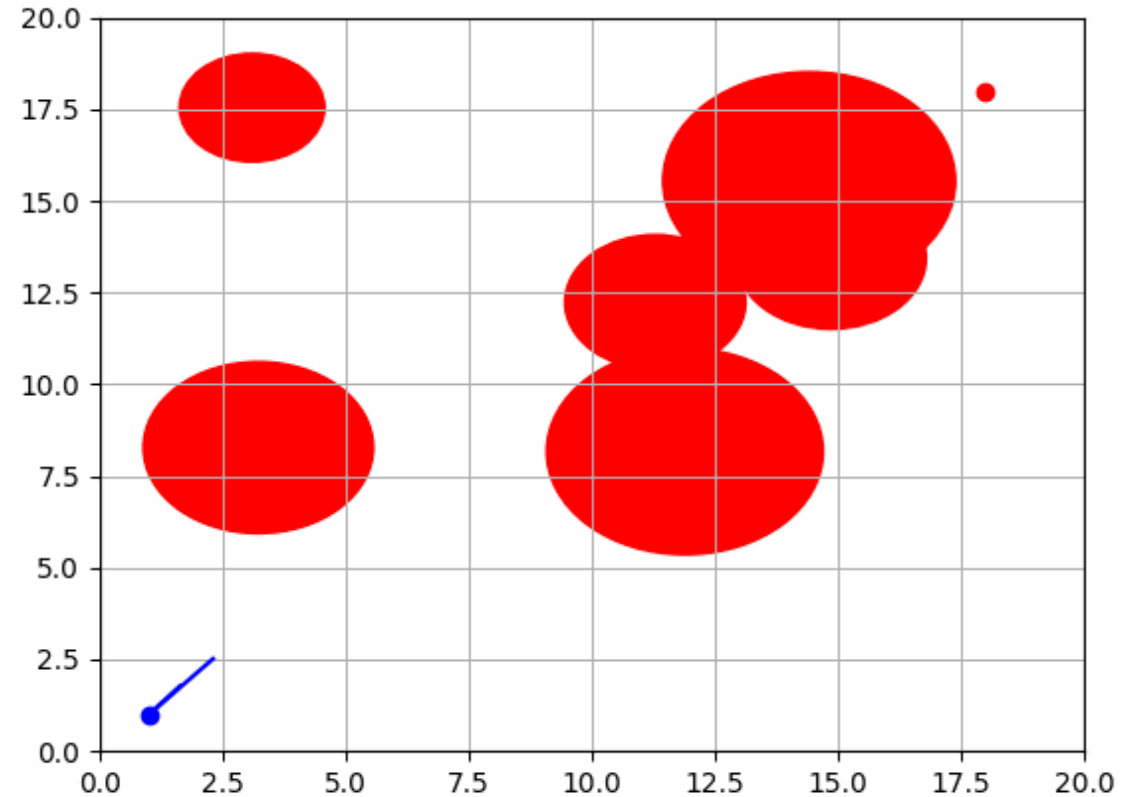
Planning

- Not enough to make individual motors spin.
- Need to coordinate motion of several motors at once.
- This is called *planning*.



Incremental Planning

- The method of the previous slide assumes we can see everything all at once.
- Is this accurate?
- When it isn't, we can “grow” a plan from where we are to where we want to go.
- This method is called ***rapidly exploring random trees (RRT)***.



Perception

- How can a robot “see” the world?
- Can't build a plan if you can't sense the world.

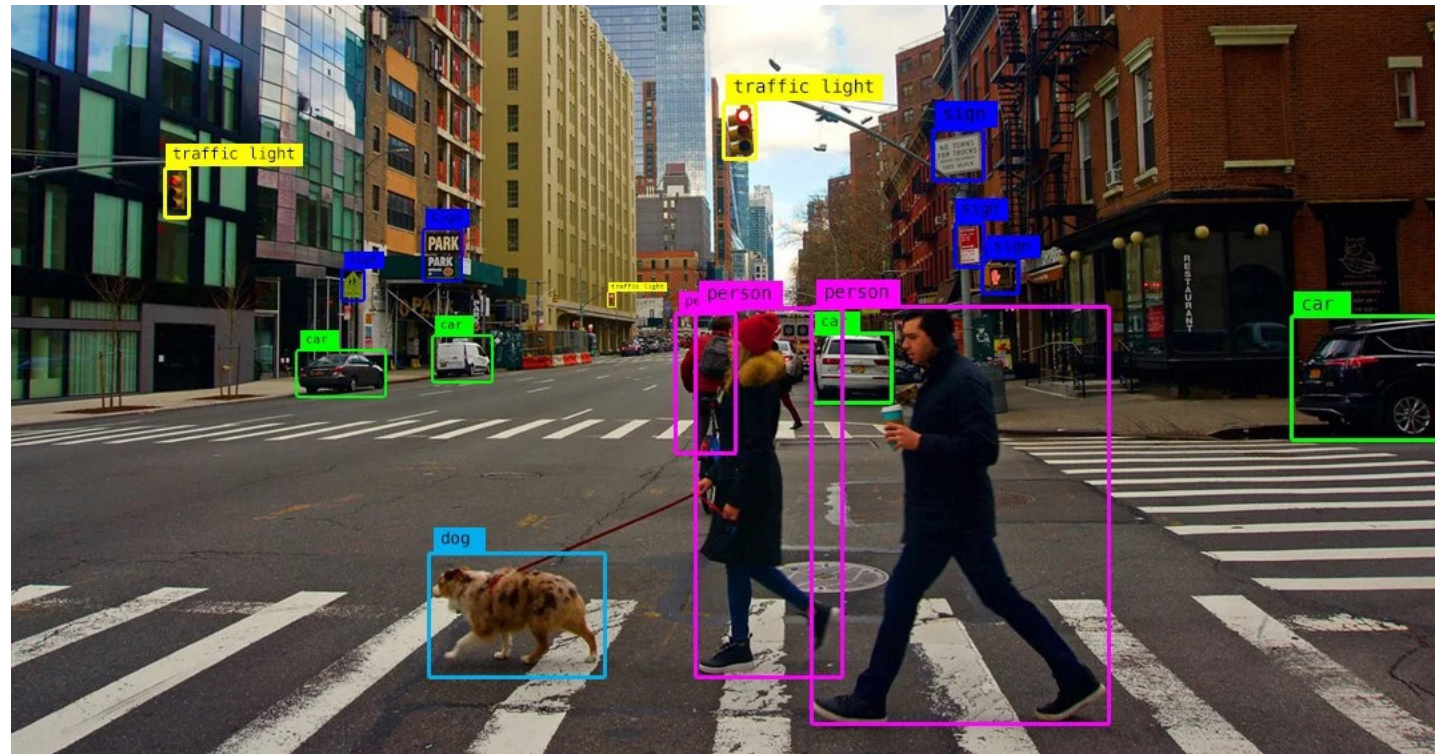
Cameras

- Cameras are now everywhere.
- There's a surprising amount of science that can be done here.
 - Enough to be a whole subject of study.



Computer Vision

- ***Computer Vision*** uses AI to process images and video.

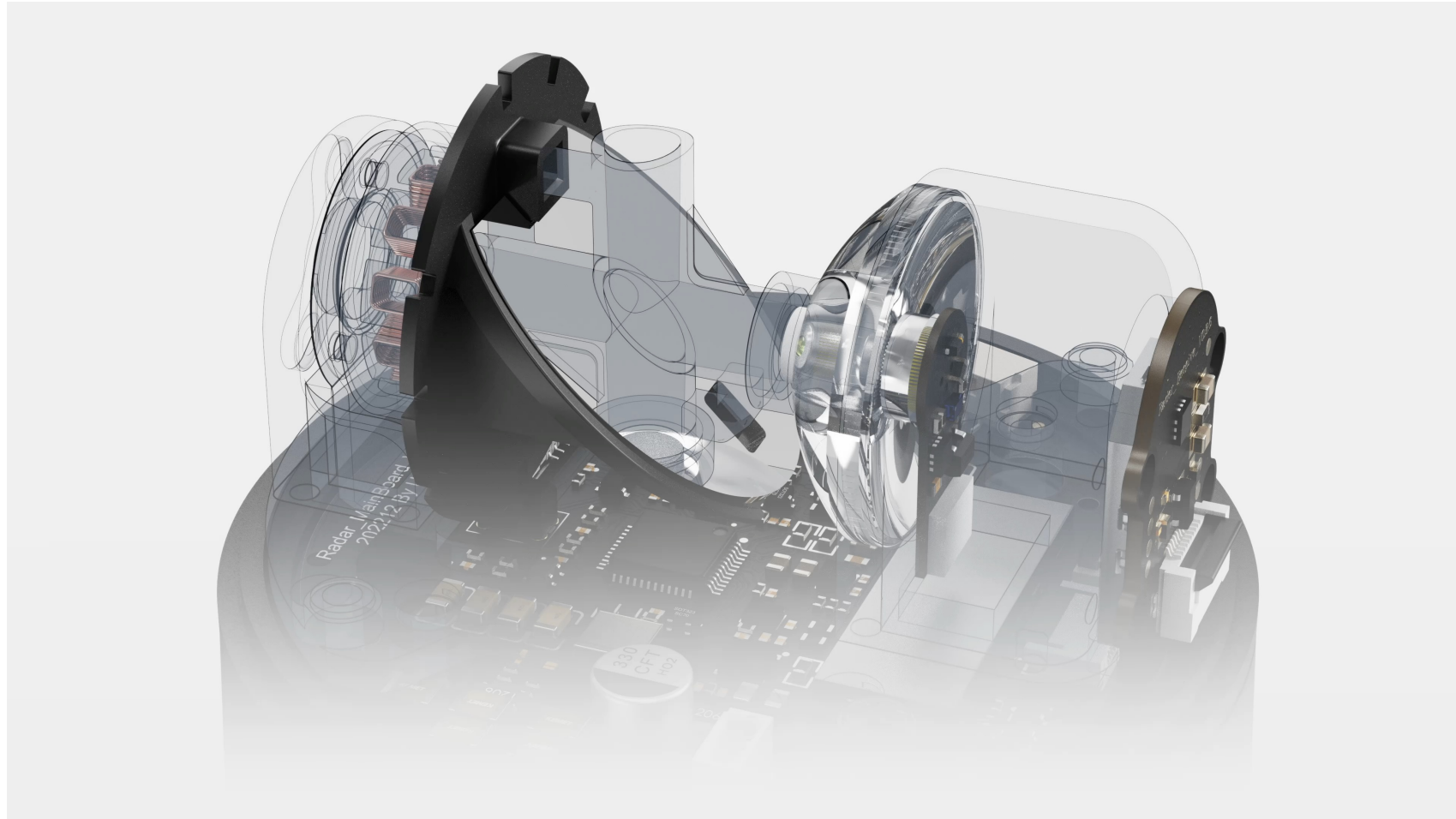


Lidar

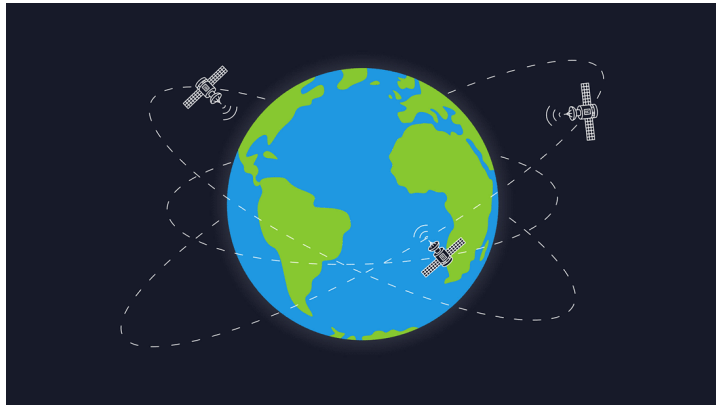
- ***Lidar*** is a key sensor in automated vehicles.
- Operation
 - Spinning platform shoots pulse of infrared light.
 - Light hits reflective object in the environment.
 - Time of flight to return to sensor gives accurate distance.
- Result: 3D ***Point Clouds***.



Quick Review: Lidar Operation

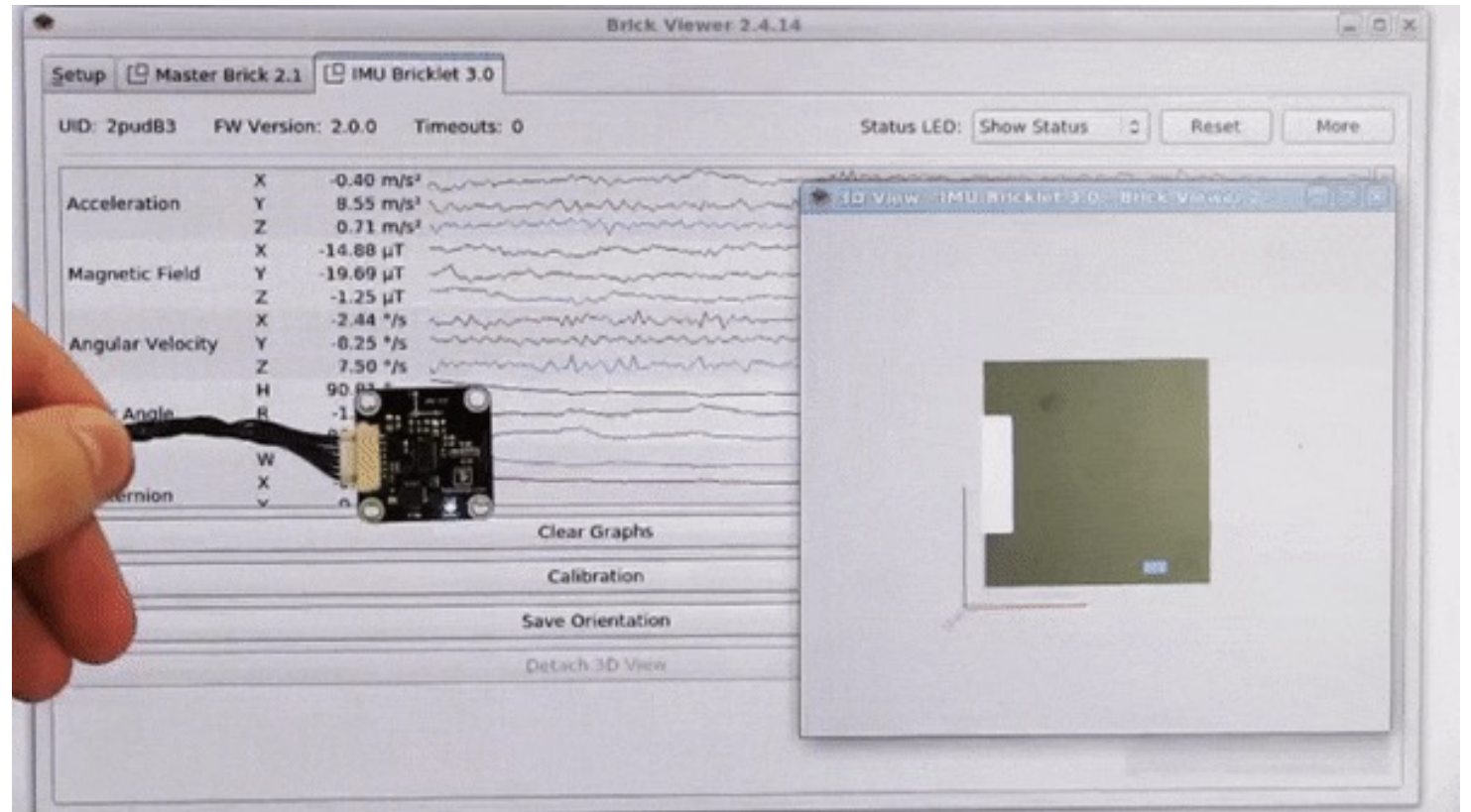


GPS & IMUs



GPS tells a robot where it is on Earth.

limitation: only works *outside*.



Inertial measurement units tell a robot how it is oriented and moving in space

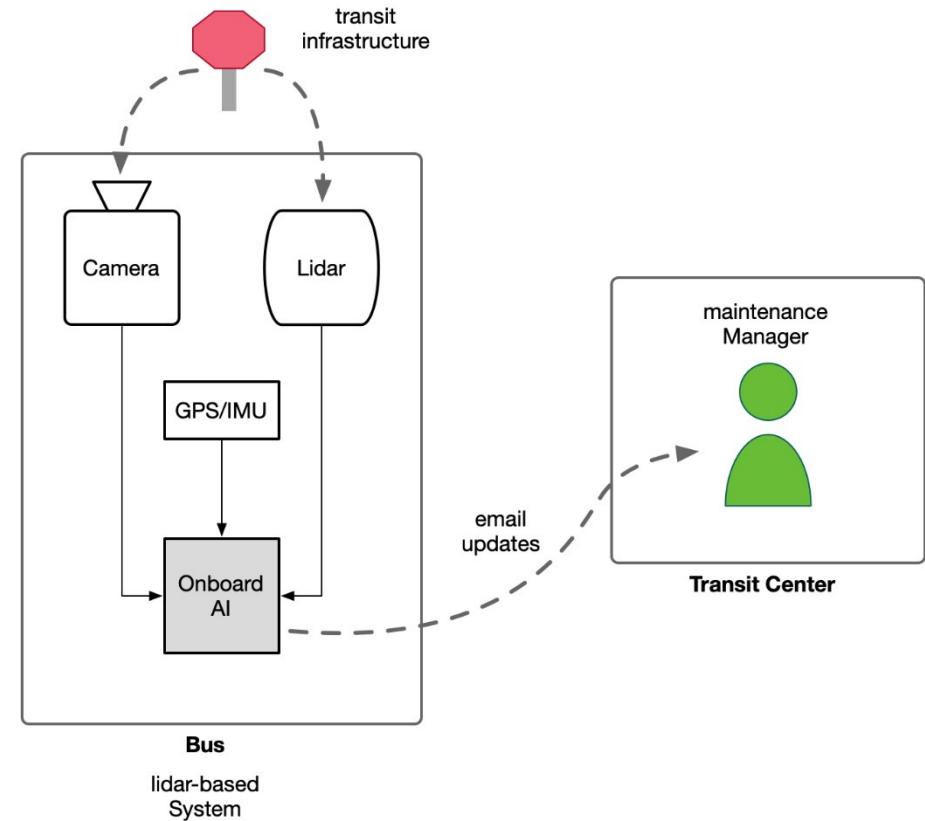
Case Study: Infrastructure Inspection

Transit Infrastructure Inspection

- In 2020, USDOT gave me money to study automated transit infrastructure inspection.
- Previous state of the art (SOTA): bus drivers look for graffiti, damaged signs, broken glass, etc. Drivers call maintenance when they get a chance.
- Hypothesis: Building a robot to automate this task would be possible (and popular with drivers and public).

Initial System Design

- Step 0: install sensors and computers on vehicles.
- Step 1: build 3D maps of bus routes using lidar.
- Step 2: detect anomalies at bus stops using cameras and lidar.
- Step 3: automatically inform maintenance staff of need for cleanup/repair/replacement...



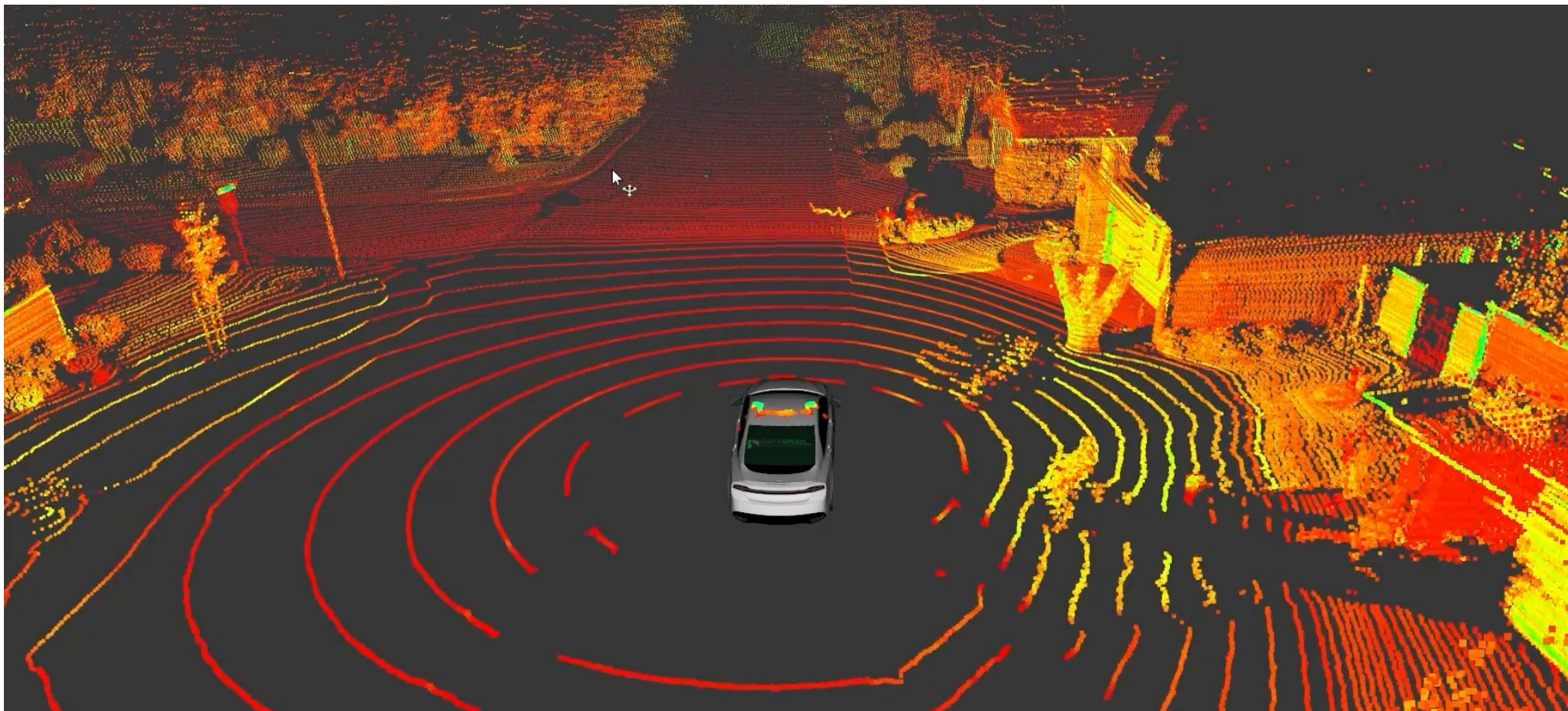
Data Collection Platforms

- A 2017 Lincoln MKZ.
 - Modified to support fully autonomous driving.
 - Supports lidar, camera connections.
- A county bus.
 - No autonomy, but modified to support sensor installation.

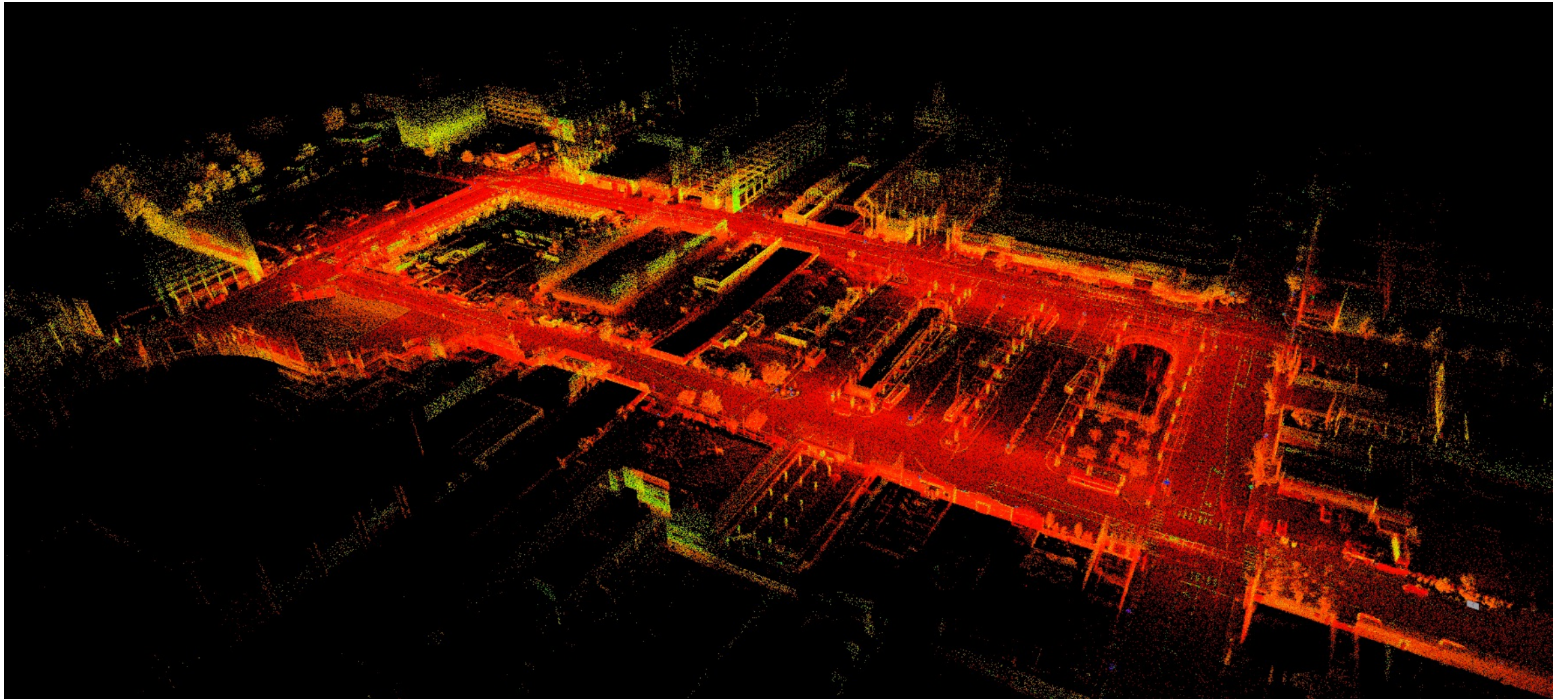




Lidar on Vehicles



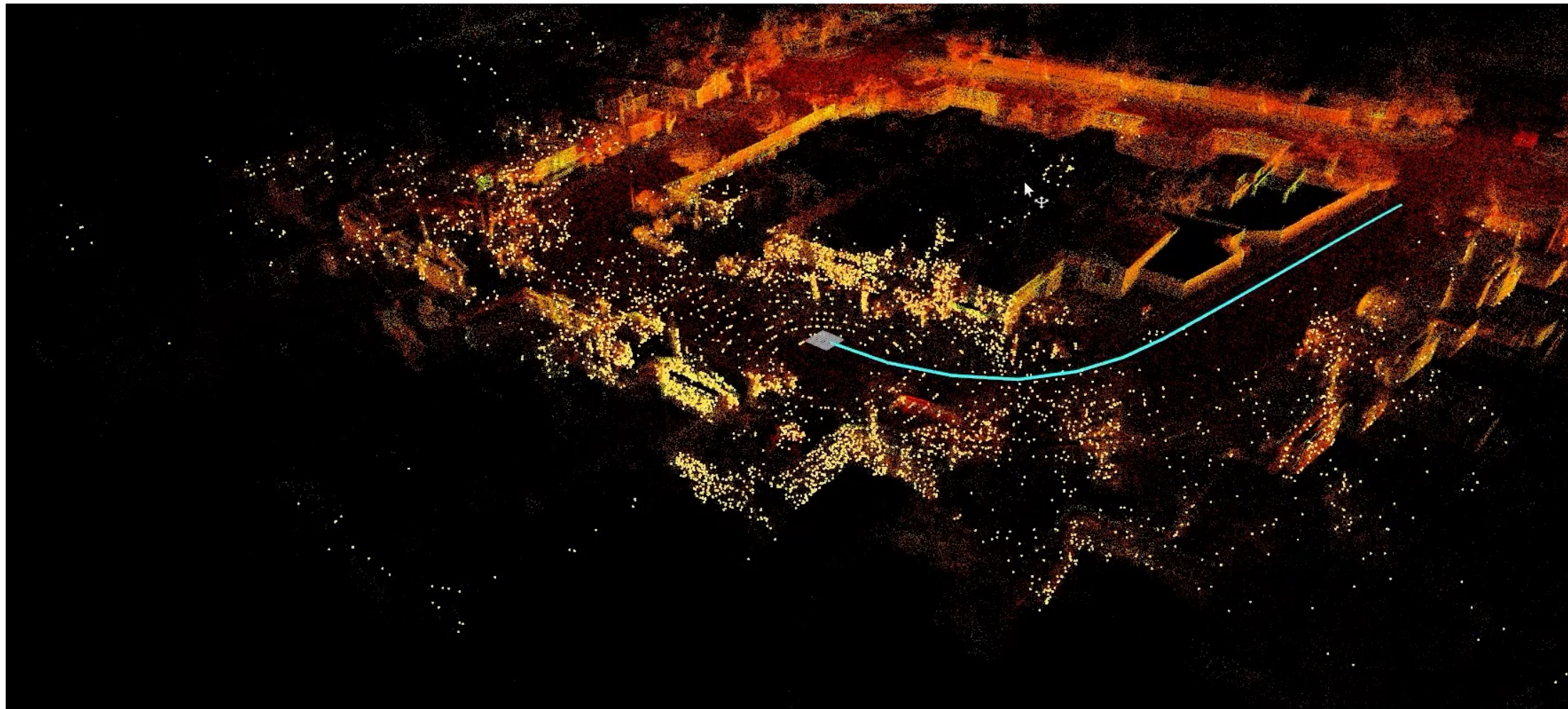
Lidar Mapping



Building a Data Pipeline



Automatic Analysis of Sites for Damage



Behavior Trees for Visual Inspection

