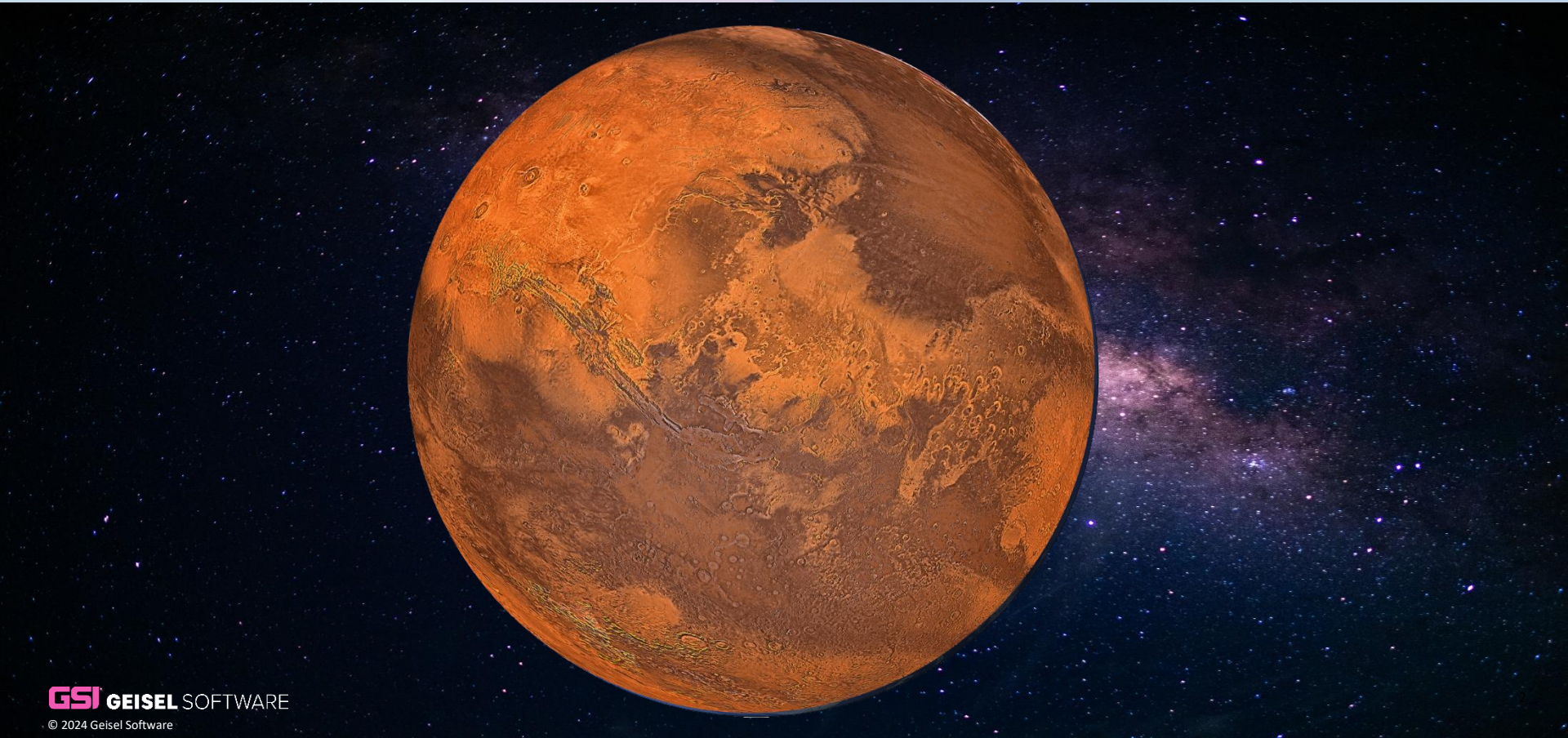
Using Synthetic Data to Train Computer Vision Models

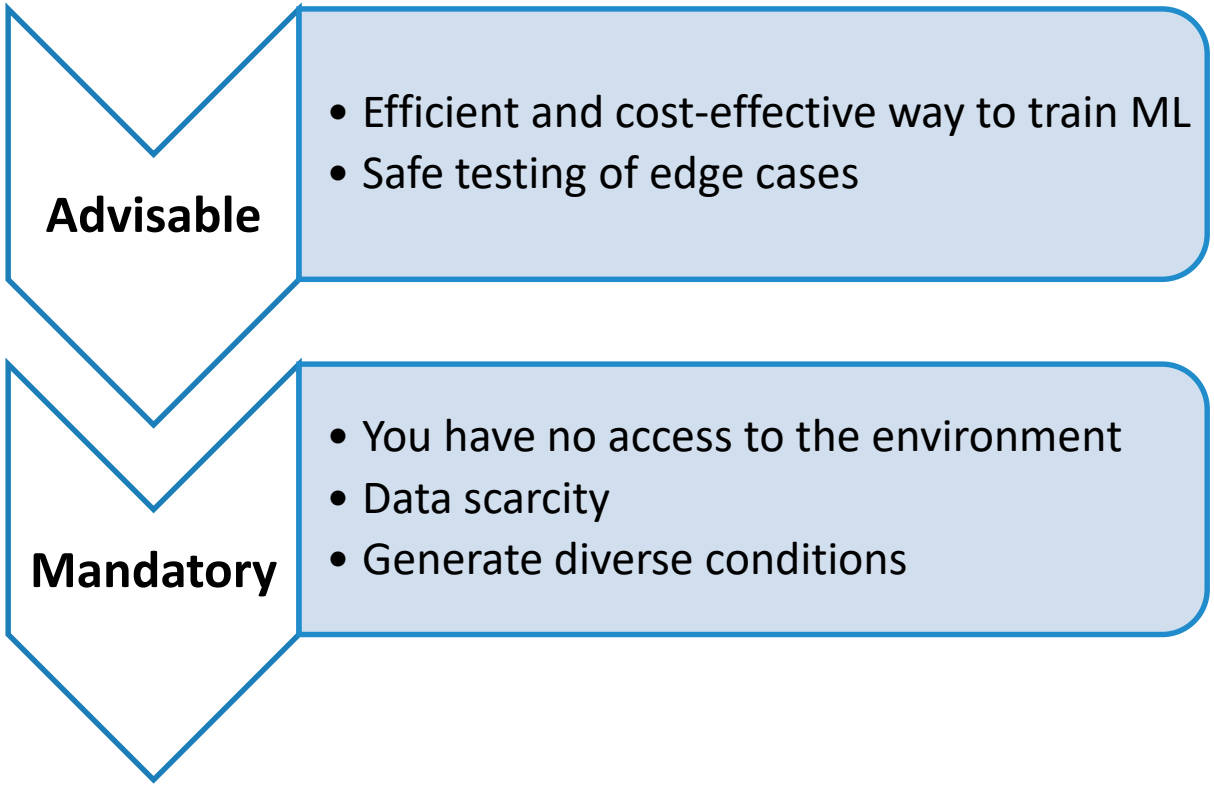
Brian Geisel, CEO
Geisel Software

GSI GEISEL SOFTWARE

Open World Simulation



Synthetic Data: When It's Advisable and When It's Mandatory



The Limitations of Synthetic Data

- Reality Gap
- Quality & Accuracy
- Overfitting
- Ethical & Legal Considerations
- Technical Complexity
- Validation Challenges

Sim2Real Gap Explained

What is the Sim2Real Gap?

The discrepancy between simulated environments and real-world conditions when developing and testing robots, algorithms, or machine learning models

Factors contributing to this gap:

- the fidelity of the physical world
- the complexity of real-world interactions
- the unpredictability of real environments versus their simplified virtual counterparts

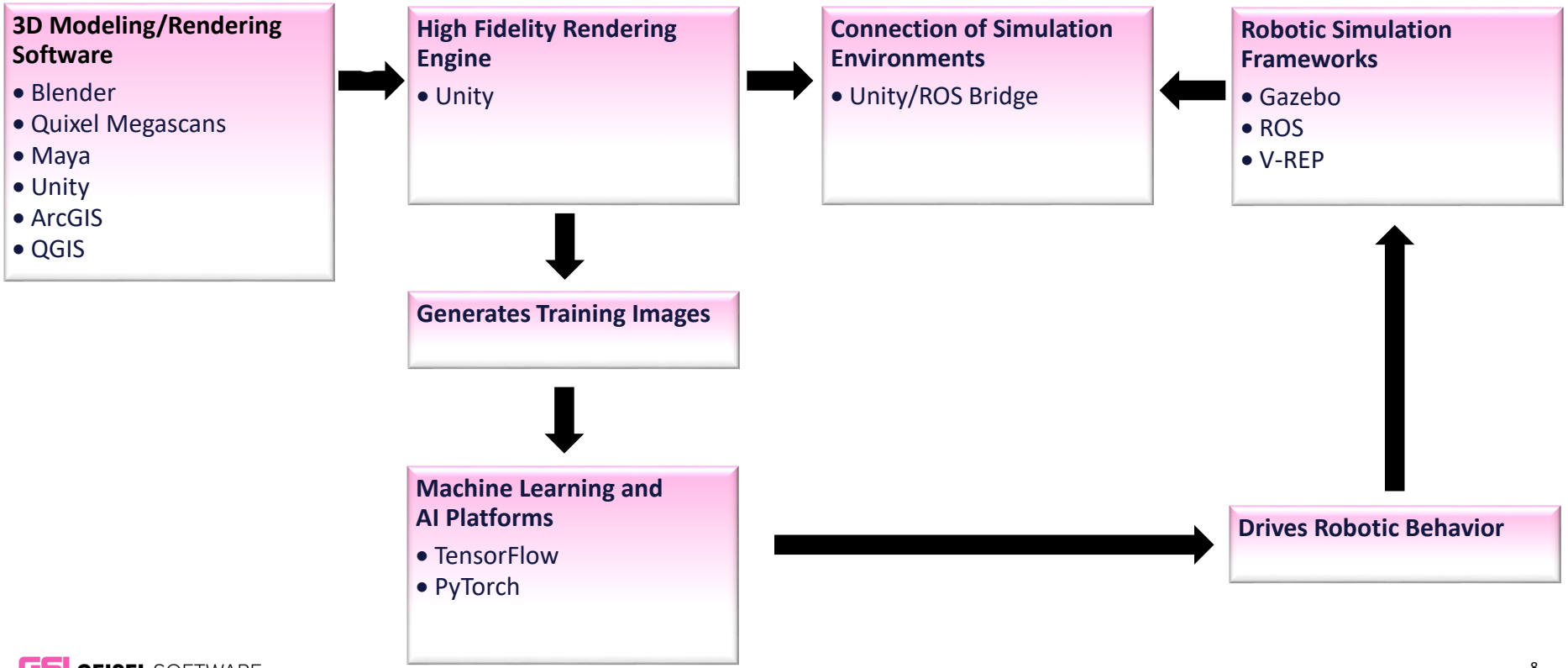
Creating Synthetic Martian Environments

The Mars Sim2Real Gap

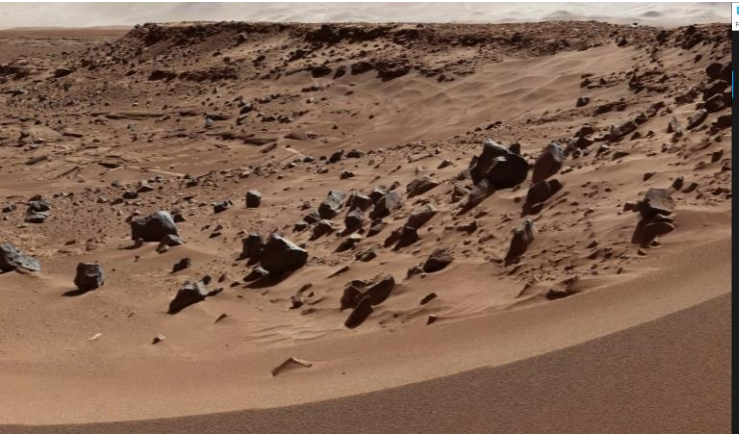
The Sim2Real gap is particularly crucial when it comes to Mars exploration:

- Unpredictable and Extreme Conditions
- Limited Training Set
- Limited Testing Opportunities
- Unable to Retrain While Deployed
- High Mission Stakes and Costs
- Limited Physical Access

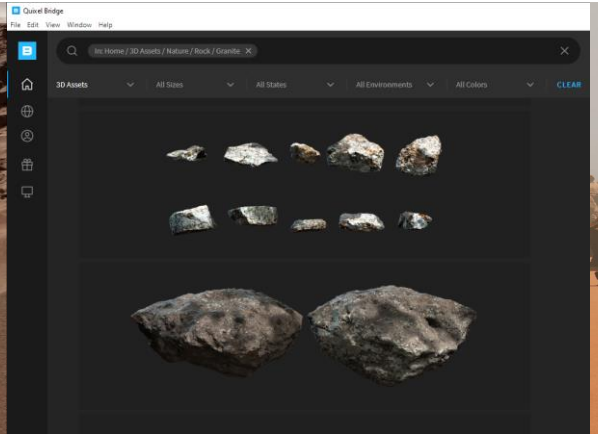
Photorealistic Simulation Tools & Technologies



Creating Training Data from Simulated Environment



Footage from Mars



Scanned Earth Objects

Create Simulation

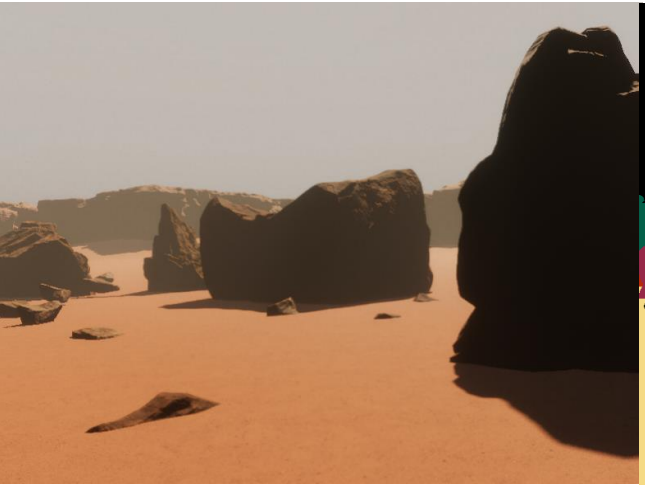
Photorealistic Simulation

Adjust Atmosphere

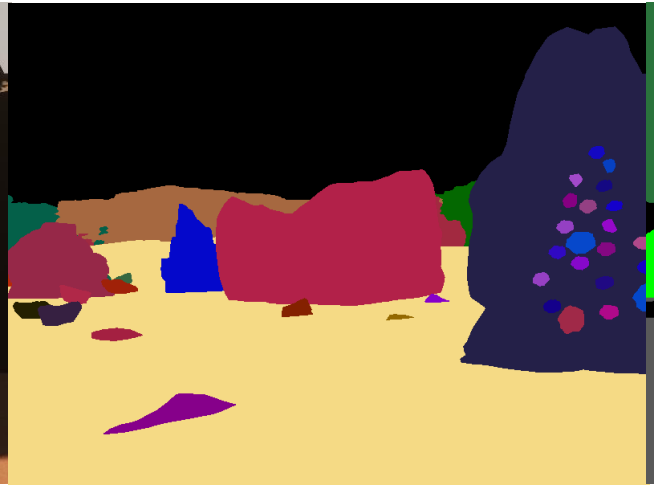
Add Noise

Shift Color

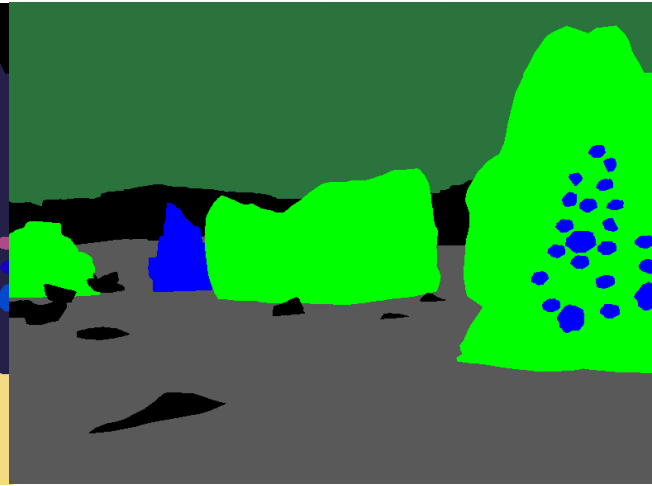
Training the Model with Synthetic Data



Original Object



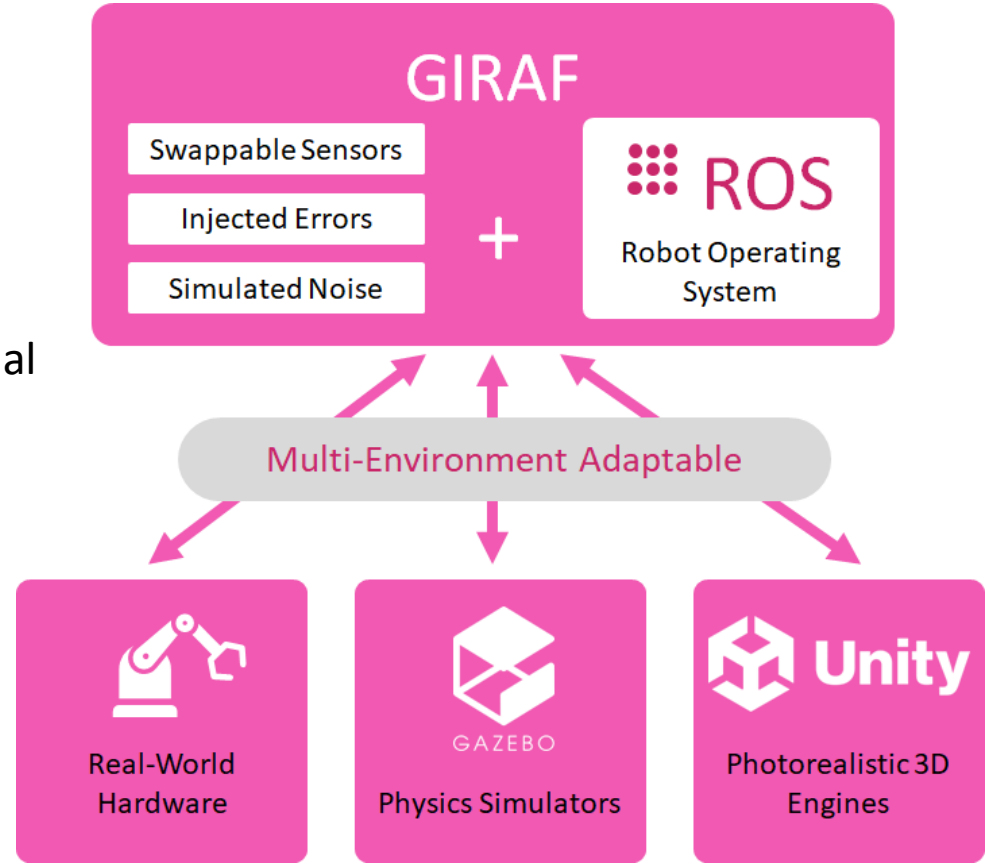
Instance Labeling



Segmentation of Objects of Interest

GIRAF RealSync Digital Twinning Platform

- Integrates digital twins of robotic systems
- Utilizes real-time data from physical systems
- Highly adaptable and requiring minimal specific system prerequisites
- Integrates seamlessly with existing robotics systems, both physical and simulated



Few-Shot Learning

Testing the Photorealism of the Simulation

Actual Atacama Desert

Resources

"Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks" by Chelsea Finn, Pieter Abbeel, and Sergey Levine
https://arxiv.org/abs/1703.03400

"Improved Synthetic Data for Deep Learning" by Lukas Tuggener, Ismail Elezi, Jürgen Schmidhuber, Thilo Stadelmann
https://arxiv.org/abs/2001.06630

Evaluation of Techniques for Sim2Real Reinforcement Learning
https://www.researchgate.net/publication/370625535_Evaluation_of_Techniques_for_Sim2Real_Reinforcement_Learning

Thanks: Let’s Connect!

Brian Geisel

https://geisel.software

brian@geisel.software

(508) 936-5099

LinkedIn

GEISEL SOFTWARE