NS&T ML-AI

Big Data *Machine Learning* *Artificial Intelligence*

This symposium involves audio/video recording of all presentations, discussions, and comments. The recording for this symposium will be made available to the public. By receiving this notification, your participation consents to recording your interactions with the symposium and public release.

Your audio and video is muted by default

Use the “Chat” feature to ask questions. All questions will be addressed at the end of each presentation (time permitting)

Use the “Chat” feature to let us know if you have technical difficulties

For low-quality connections, switch off video and do not use VPN, if possible

A separate audio PIN will be provided when you sign in for the phone-in audio option

Webinar will begin at 11:00 am MST

Welcome to the

Artificial Intelligence and Machine Learning Symposium10.0

January 19, 2023

IDAHO NATIONAL LABORATORY

January 19th, 2023

Dr. Curtis Smith, Director
Nuclear Safety and Regulatory
Research Division

Welcome to the AI/ML Symposium 10.0

NS&T ML-AI

Big Data, Machine Learning,
Artificial Intelligence

"AI Technology"

Agenda – ML/AI Symposium 10.0

January 19, 2023 - 11:00 am to 1:00 pm MDT

Time	Presentation Subject	Speaker(s)
11:00 – 11:10	Welcome, Introductions, and Agenda	Curtis Smith Director, Nuclear Safety and Regulatory Research Division INL
11:10 - 11:35	Creative AI: Pop* and Therapeutic Computational Creativity	Paul Bodily Assistant Professor, Computer Science Idaho State University
11:35 – 12:00	Accelerated Data Science and AI	Patty Delafuente Data Scientist, Solution Architect NVIDIA
12:00 – 12:25	AMD Instinct & AI	Ian Ferreira Sr Director, AI Software Solutions AMD
12:25 – 12:50	Tentacle-Like Continuum Robots and Sub-Task Automation	Alan Kuntz Professor University of Utah
12:50 -1:00	Closeout	Curtis Smith Director INL

Idaho State University

Creative AI: Pop* and Therapeutic Computational Creativity

Paul Bodily

Assistant Professor

Department of Computer Science

Idaho State University

ROAR

Computational Creativity

- What is Computational Creativity?

*"The philosophy, science and engineering of **computational systems** which, by taking on particular responsibilities, exhibit behaviors **that unbiased observers would deem to be creative**."*

Simon Colton and Geraint A. Wiggins. Computational creativity: The final frontier? In *Proceedings of the Twentieth European Conference on Artificial Intelligence*, pages 21–26. IOS Press, 2012.

What is *creativity*?

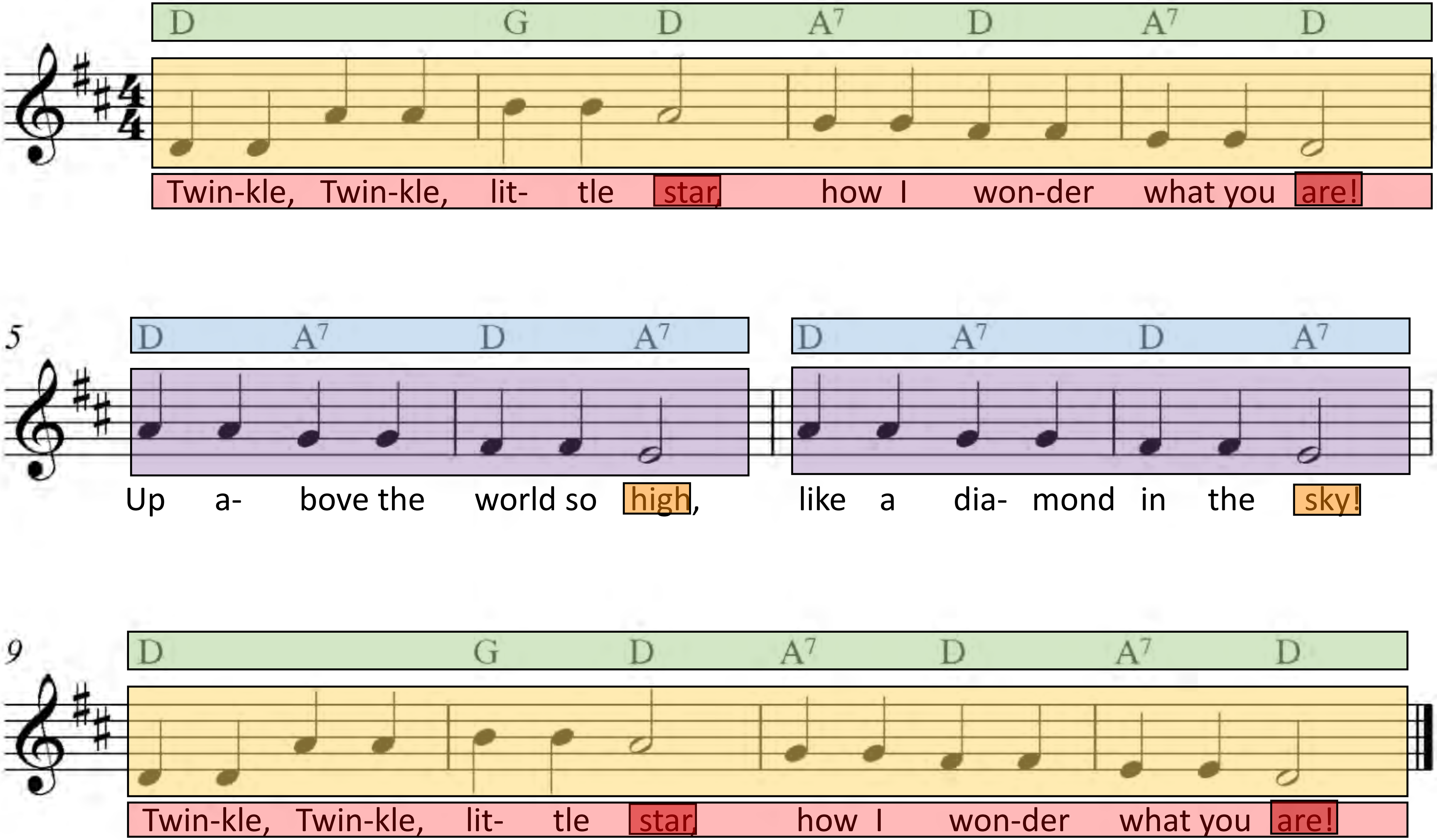
How can **computers** be *creative*?

How can we evaluate creative systems?

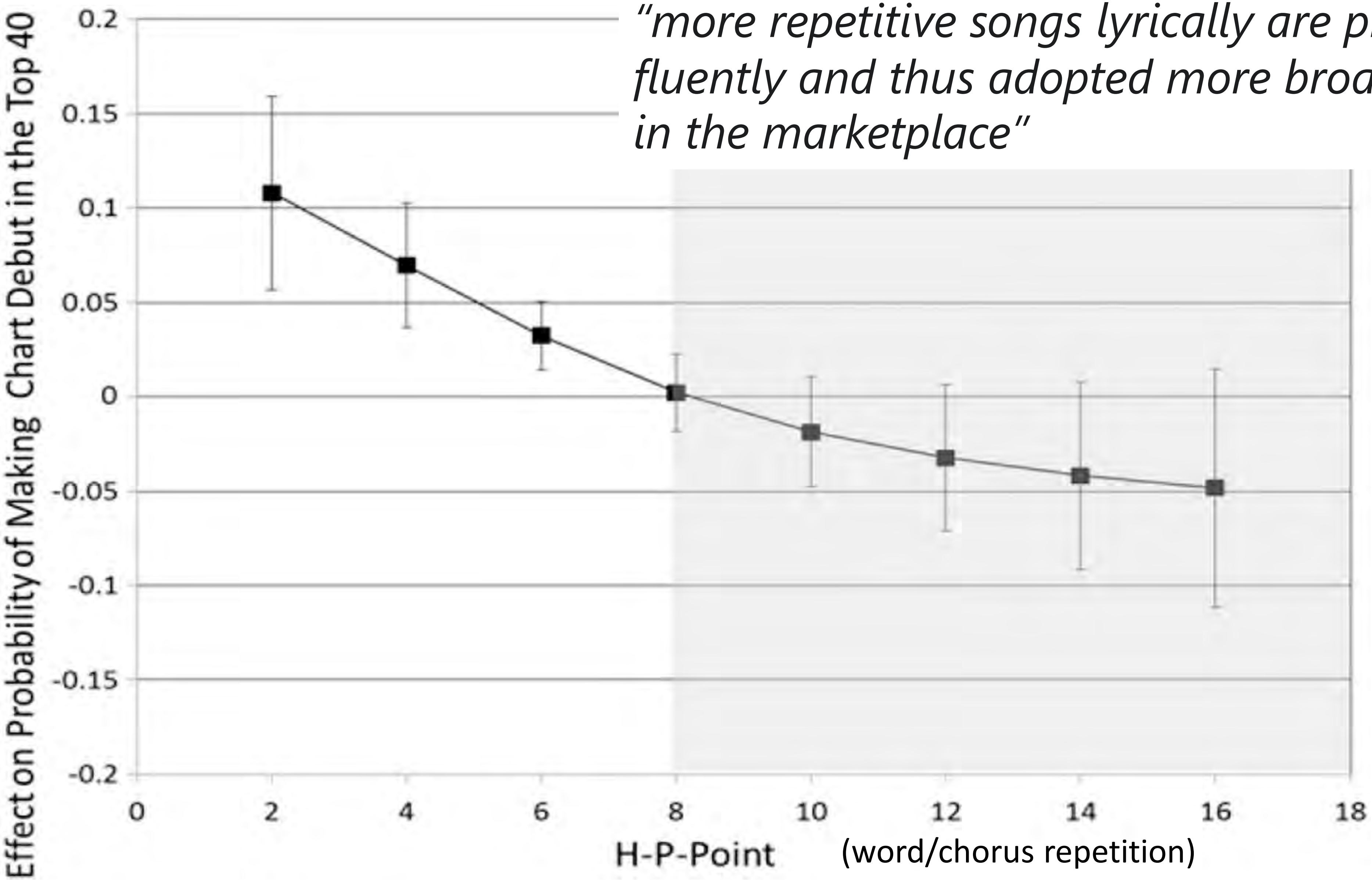
What defines creativity?

Graeme Ritchie. Some empirical criteria for attributing creativity to a computer program. *Minds and Machines*, 17(1):67–99, 2007.

Twinkle, Twinkle, Little Star



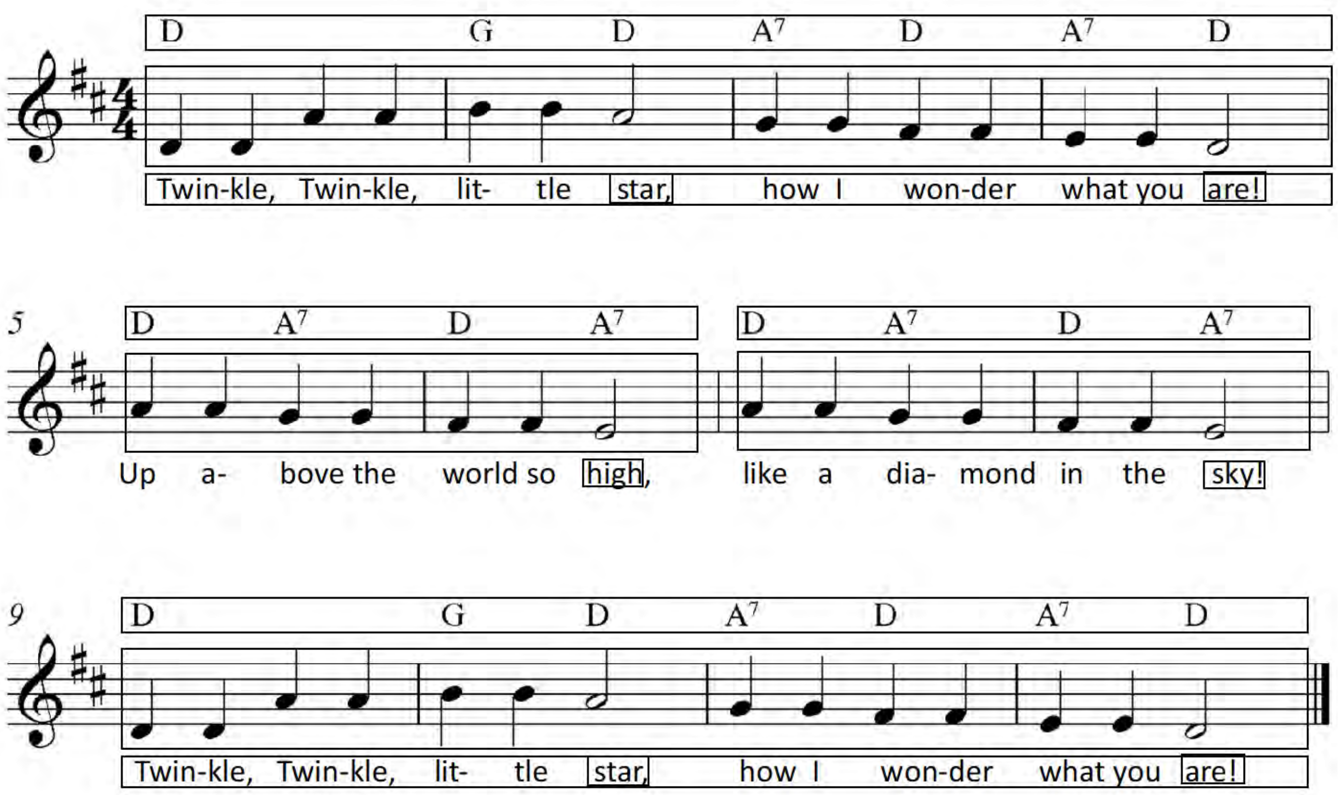
“more repetitive songs lyrically are processed more fluently and thus adopted more broadly and quickly in the marketplace”



(word/chorus repetition)

Joseph C. Nunes, Andrea Ordanini, and Francesca Valsesia. The power of repetition: Repetitive lyrics in a song increase processing fluency and drive market success. *Journal of Consumer Psychology*, 25(2):187–199, 2014.

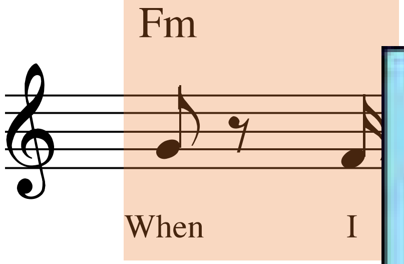
Problem: Long-Term Dependency Challenge

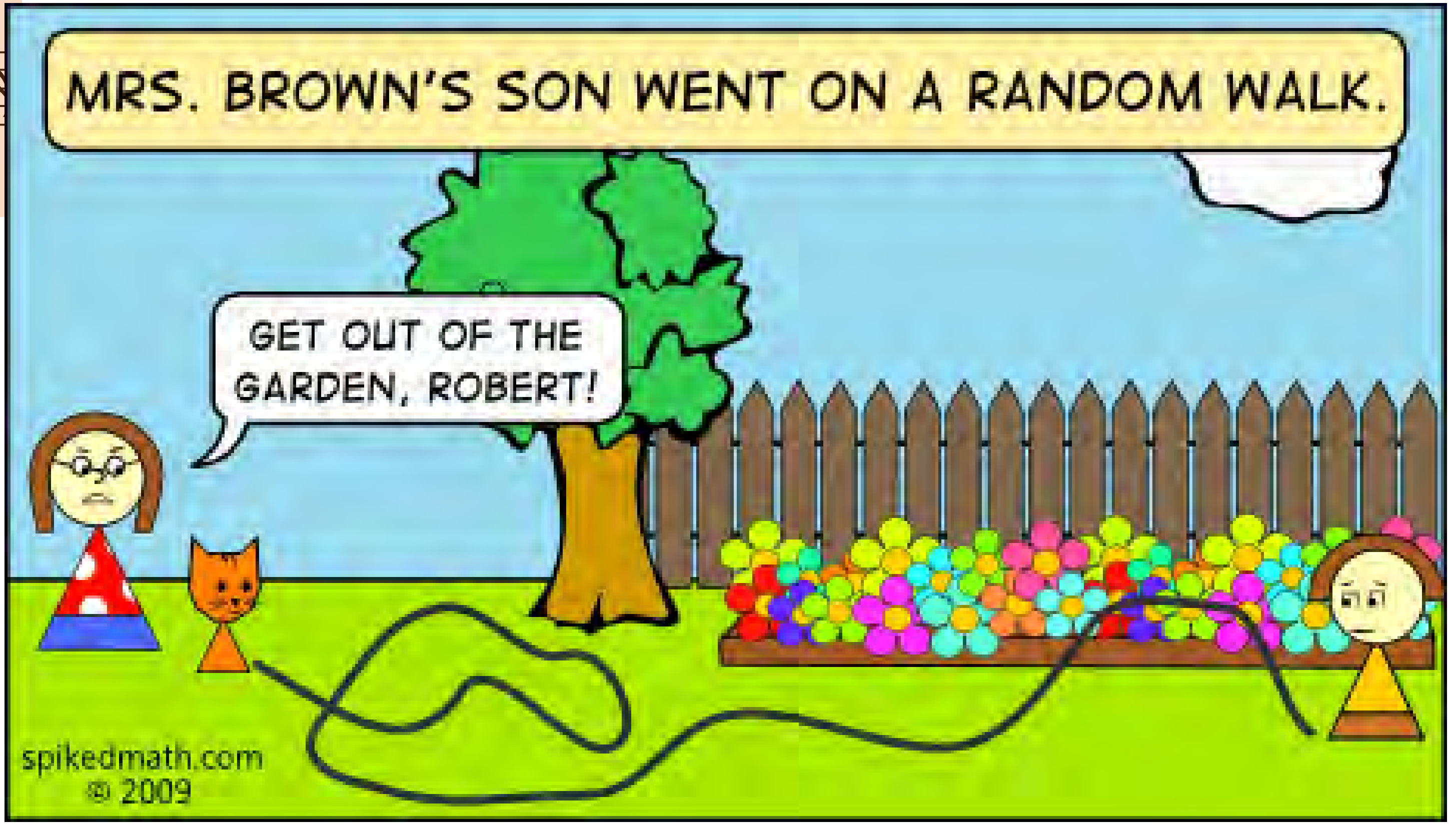


Almost all models "are either of a Markovian kind, assuming a strictly **limited range of dependency** of the musical present on the musical past … or have a kind of **decaying memory** (as in simple Recurrent Neural Networks). On the other hand, it is clear that music is of a fundamentally non-Markovian nature. **Music is full of long-term dependencies**: … themes return at regular or irregular intervals, after some intermittent material … we need more research on complex temporal models with variable degrees of memory" (p. 4).

Widmer, Gerhard. "Getting closer to the essence of music: The Con Espressione manifesto." ACM Transactions on Intelligent Systems and Technology (TIST) 8.2 (2017): 19.

Problem: Long-Term Dependency Challenge





Local Structure

Global Structure

Problem: Long-Term Dependency Challenge

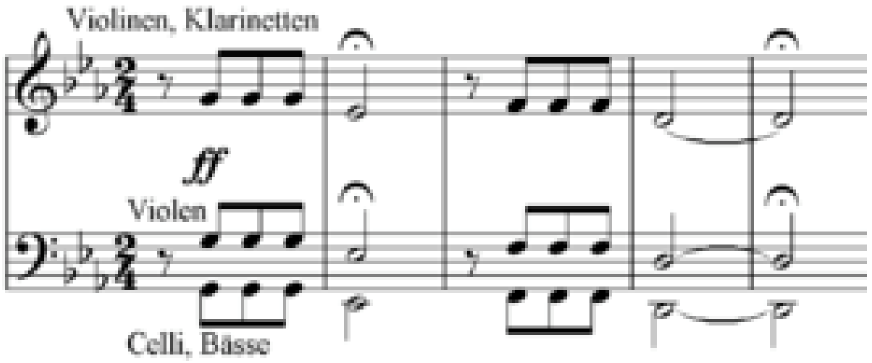
?

 Local Structure

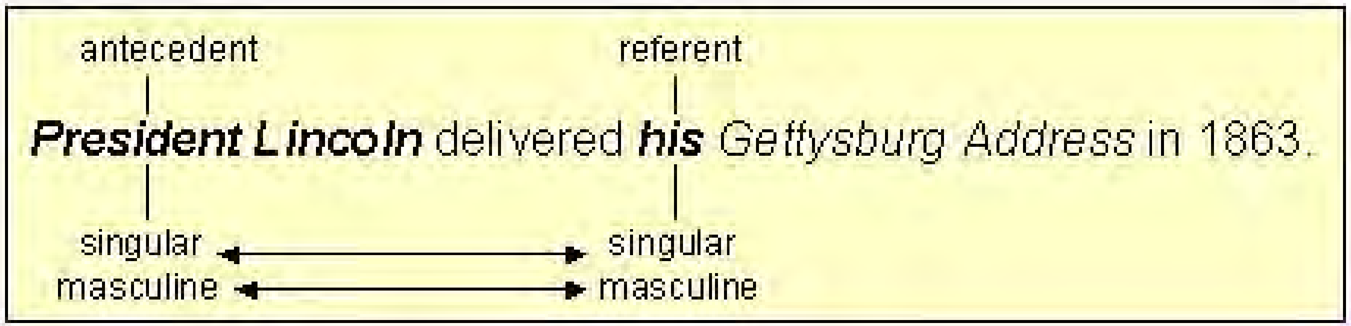
 Global Structure

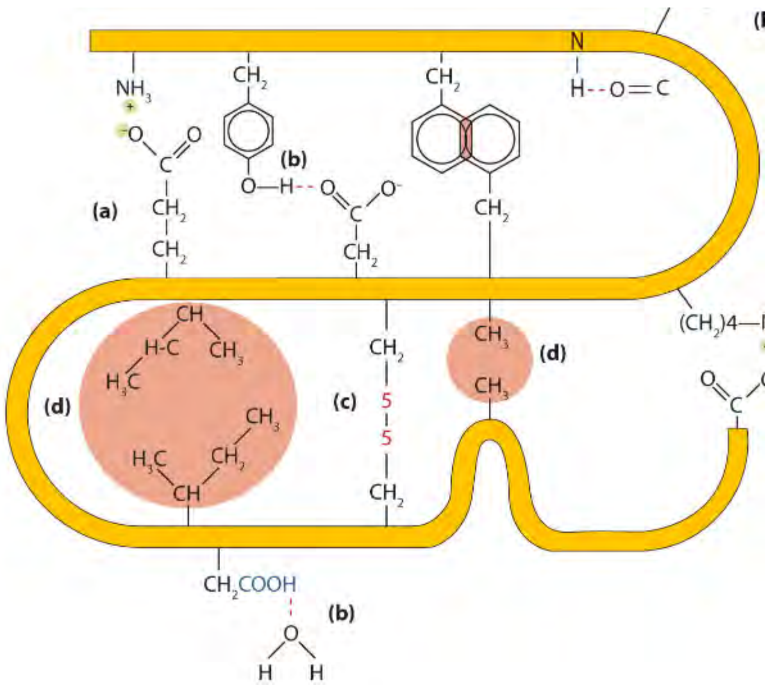
Problem Statement

There is a general lack of models capable of learning and exhibiting both local cohesion and relational global structure in sequence generation.



There is a note in his eyes
He backs the beat of the key
Down the song in my eyes
You back the beat of the sea

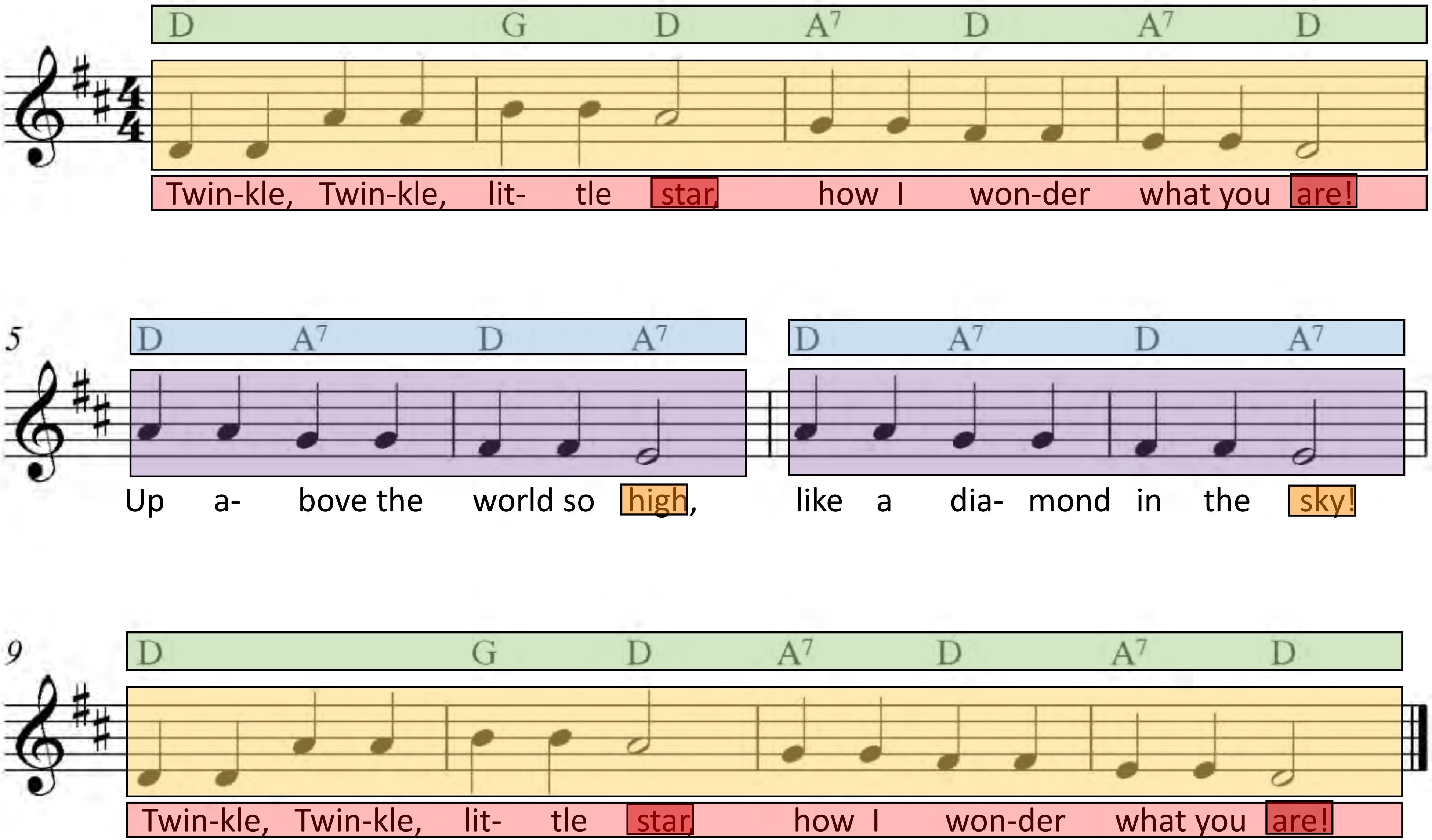


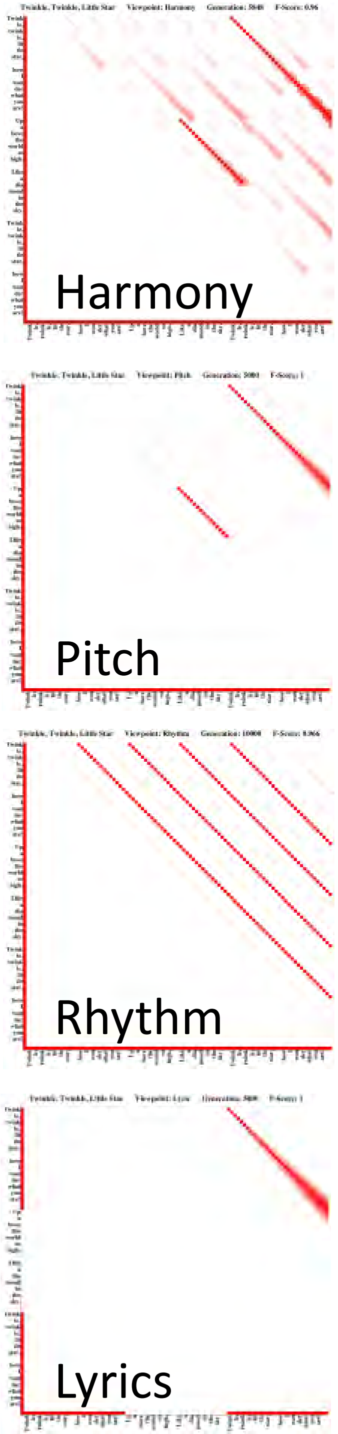


Challenge #1

Learning to recognize long-term dependencies in existing music

Twinkle, Twinkle, Little Star



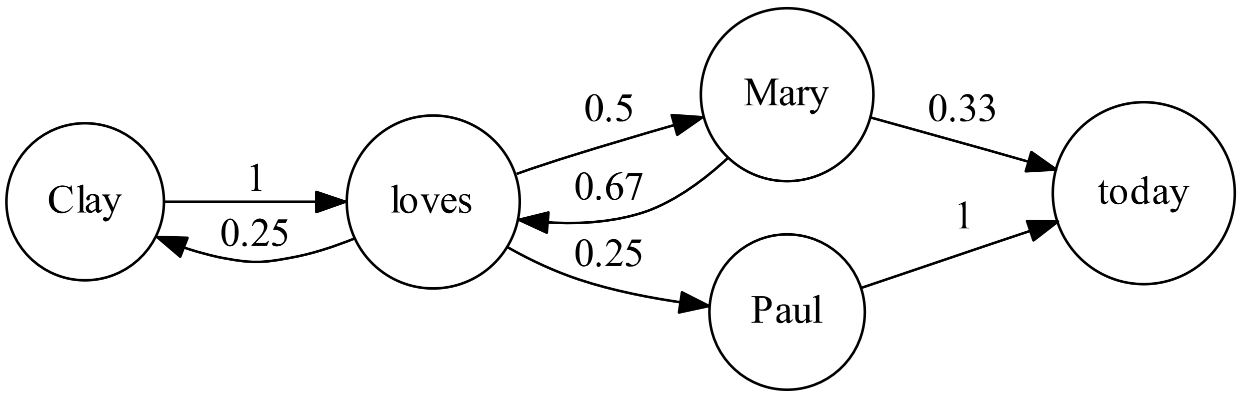


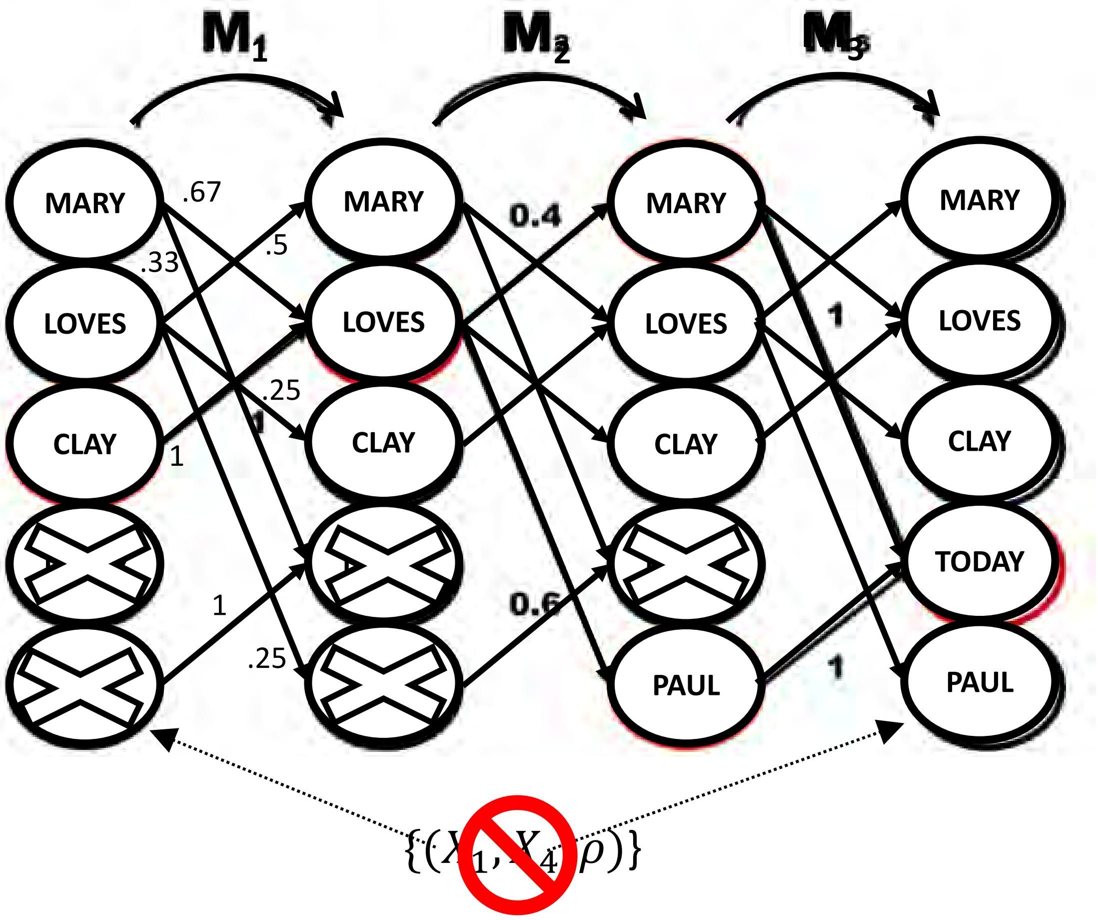
Challenge #2

Creating a model to reproduce long-term dependencies

Constrained Markov models

- *Clay loves Mary*
- *Mary loves Clay*
- *Clay loves Mary today*
- *Mary loves Paul today*

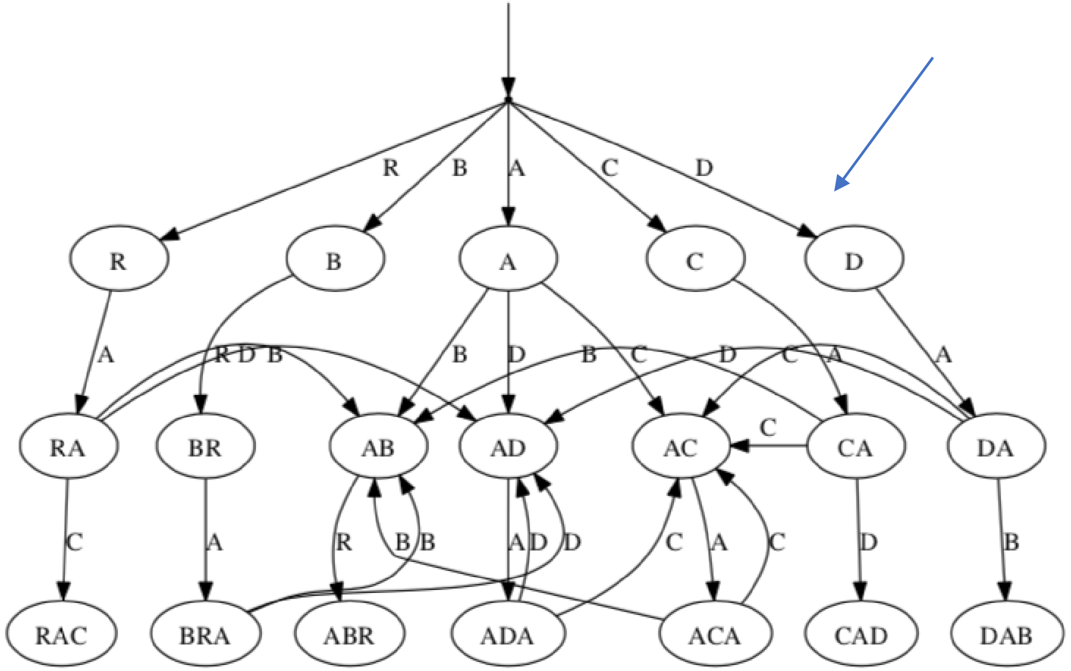




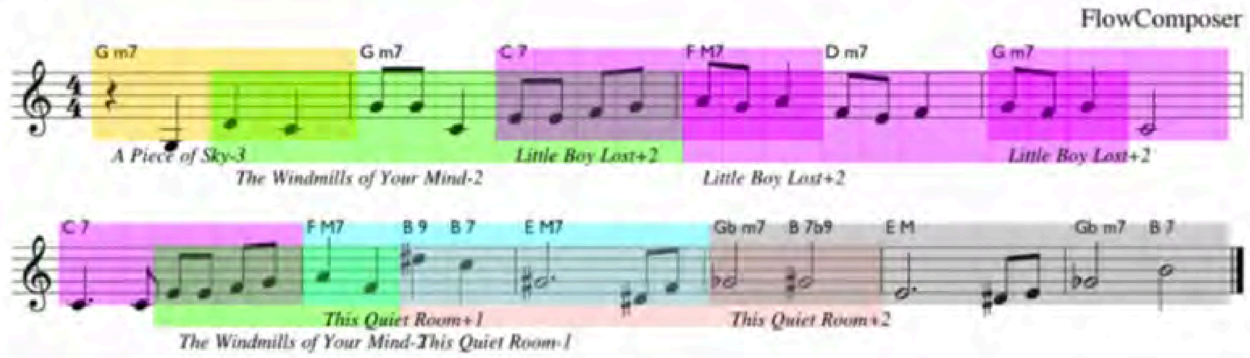
Gabriele Barbieri, François Pachet, Pierre Roy, and Mirko Degli Esposti. Markov constraints for generating lyrics with style. In *Proceedings of the Twentieth European Conference on Artificial Intelligence*, pages 115–120. IOS Press, 2012.

Imposing structure with DFAs

Deterministic Finite Automaton



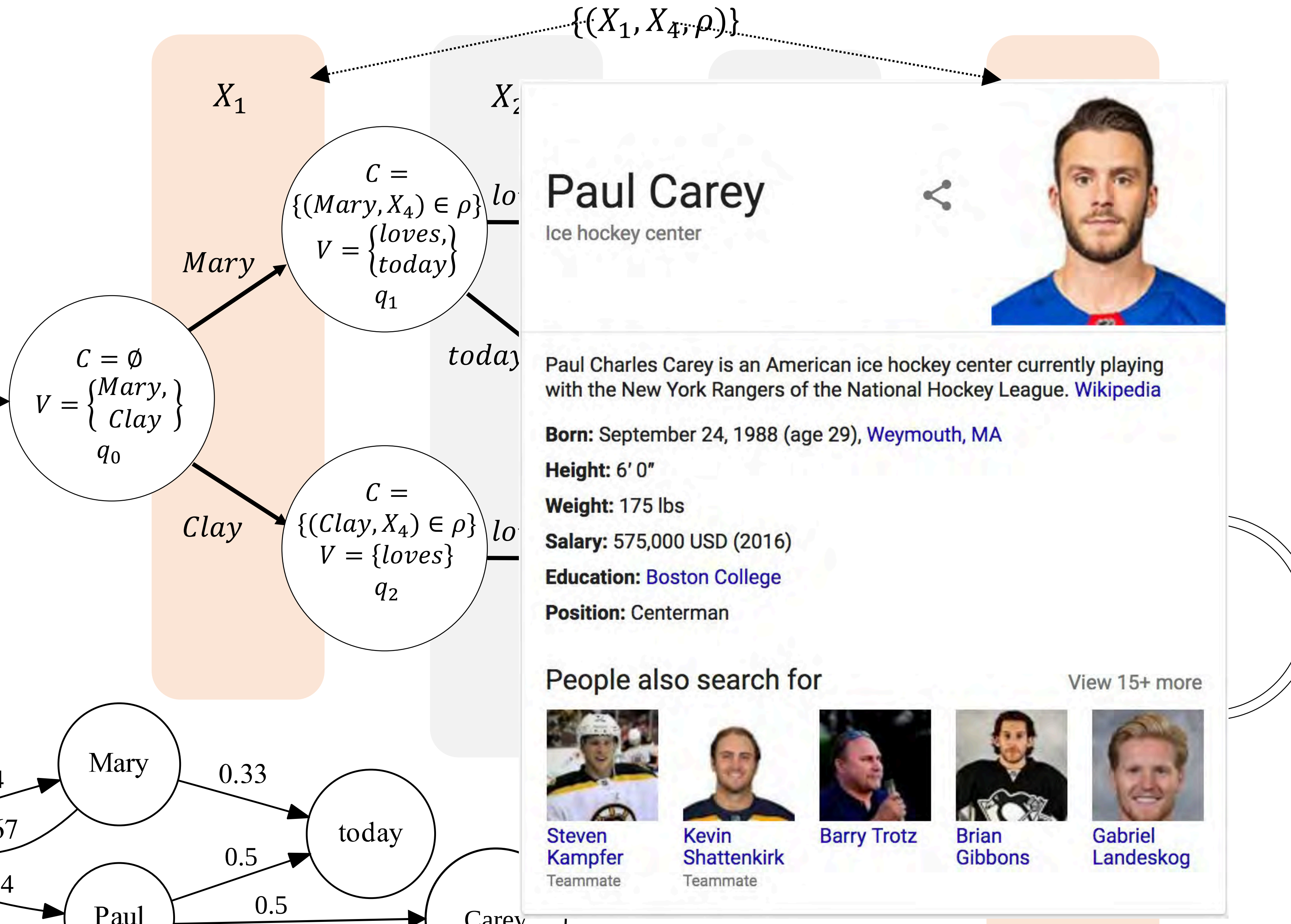
DFA that accepts all subsequences of ABRACADABRA of length 4 used to constrain against plagiarism in Markov processes.

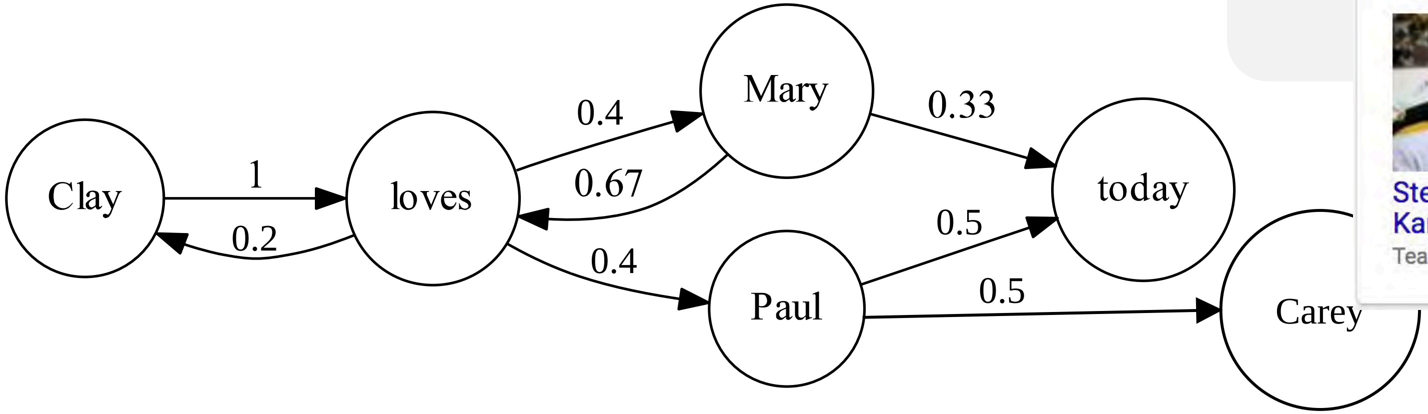


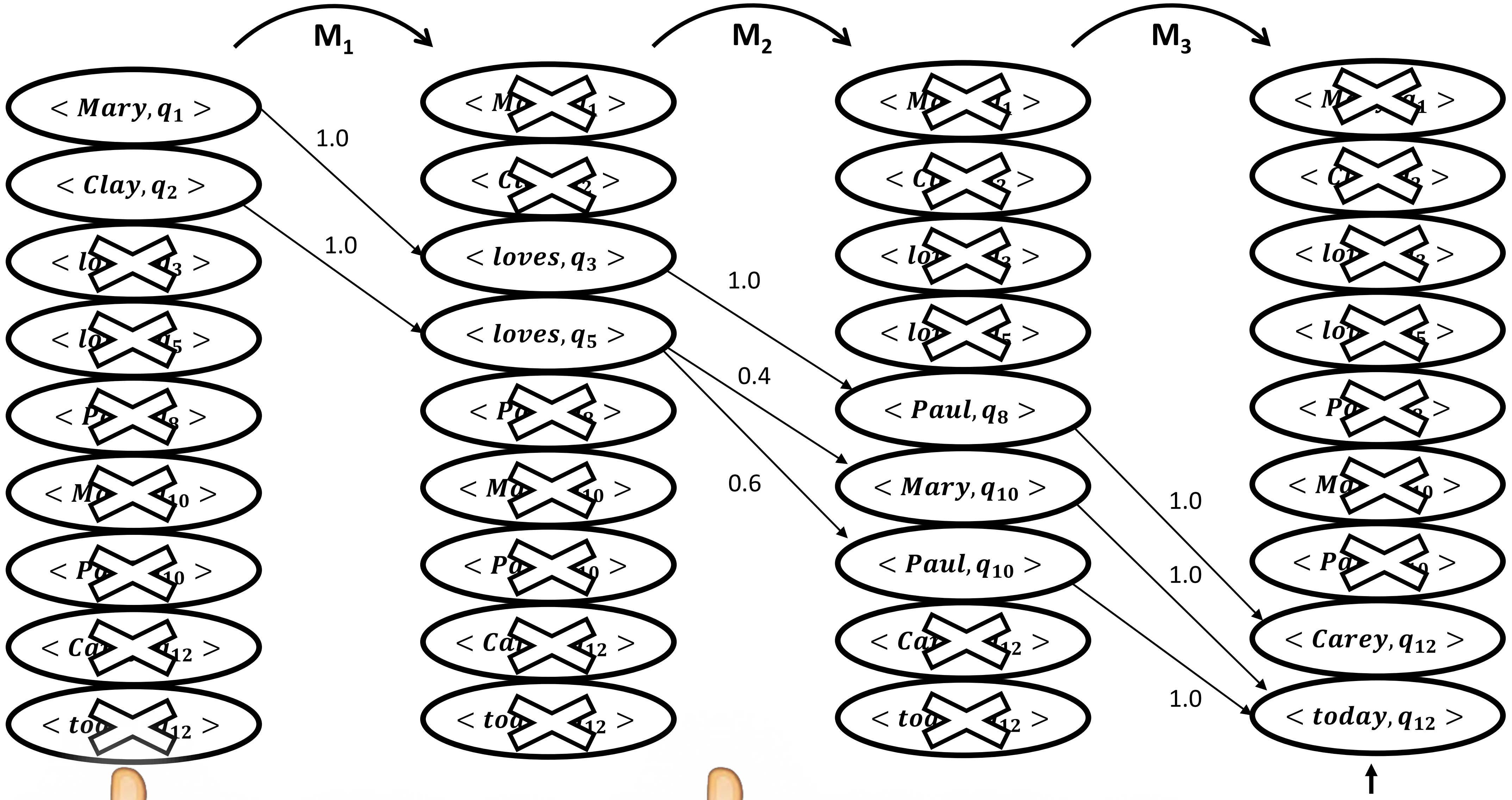
A Lead sheet generated which contains no plagiarized subsequences of length >= 6.

Alexandre Papadopoulos, Pierre Roy, and François Pachet. Avoiding plagiarism in Markov sequence generation. In *Proceedings of the Association for the Advancement of Artificial Intelligence Conference on Artificial Intelligence*, pages 2731–2737, 2014.

- *Clay loves Mary*
- *Mary loves Clay*
- *Clay loves Mary today*
- *Mary loves Paul today*
- *Clay loves Paul Carey*







Local Structure Global Structure

$\mathbf{C}_n$

q_i must be an accepting state

Challenge #3

Applying trained models with constraints in a creative AI system

Idaho State University

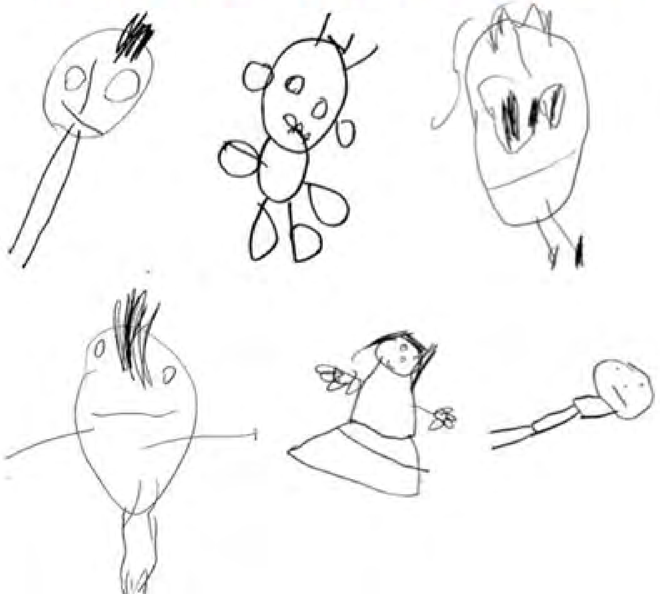
How a computer *learns:* From examples

- *"The key is not the* **stuff** *out of which brains are made, but the* **patterns** *that can come to exist inside the stuff of a brain." (Hofstadter)*
- Volunteer:
  - *Think of* an *adjective* or *concept* (e.g., green, hot, beautiful, etc.), *but don't say what it is!*
  - *Score each of the following images on a scale of 1 to 10 of how well the image embodies the concept*
- Everyone else:
  - *Looking at the scores, what do you think the concept is?*

ROAR

Idaho State University

How a computer *learns:* From examples



Idaho State University

How a computer *learns:* From examples

- The **volunteer** represents the **context** in which creativity occurs (discipline, society, universe)
  - *Concepts exist within the society, independent of the learner*
  - *Various artifacts are deemed by the discipline or society to be more or less representative of a concept*
- **Everyone else** represents the **learner** (the creative agent)
  - *Abstracts patterns representative of the concept from examples which are then replicated with appropriate diversification*
  - *Does it matter if we can put a name to the concept?*
- What **concepts** is a computer learning when given a knowledgebase, e.g., the Beatles?
  - *"good music", "Beatlesesque", "creative", "well-structured music", "non-random", "listenable"*

Harmony Pitch Rhythm Lyric

```
Prob:0.02501208850264606, XProb:0.04952        CM7 Am7 Em7 Em7 Dm7 G7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 Dm7 Bb7 CM7 Am7 Em7 Em7 Dm7 G7 CM7 CM7
Prob:2.76909019216176lE-4, XProb:3.5E-4 CM7 CM7 CM7 CM7 Dm7 Dm7 CM7 CM7 CM7 CM7 Dm7 G7 CM7 CM7 Dm7 G7 CM7 CM7 CM7 CM7 Dm7 Dm7 CM7 CM7
Prob:0.001190658097202680 4, XProb:0.00237     CM7 Em7 Dm7 Bb7 CM7 CM7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 Dm7 Bb7 CM7 Em7 Dm7 Bb7 CM7 CM7 CM7 CM7
Prob:4.940659456842380 7E-5, XProb:1.1E-4      CM7 Am7 Em7 A7 Dm7 Dm7 CM7 CM7 CM7 Em7 Dm7 Dm7 CM7 Em7 Dm7 Dm7 CM7 Am7 Em7 A7 Dm7 Dm7 CM7 CM7
Prob:2.769090192161751 3E-4, XProb:1.9E-4      CM7 CM7 Dm7 G7 CM7 CM7 CM7 CM7 CM7 CM7 Dm7 Dm7 CM7 CM7 Dm7 Dm7 CM7 CM7 Dm7 G7 CM7 CM7 CM7 CM7
Prob:0.01600773664016933 6, XProb:0.06333      CM7 Am7 Em7 Em7 Dm7 Bb7 CM7 CM7 CM7 Em7 Dm7 G7 CM7 Em7 Dm7 G7 CM7 Am7 Em7 Em7 Dm7 Bb7 CM7 CM7
```

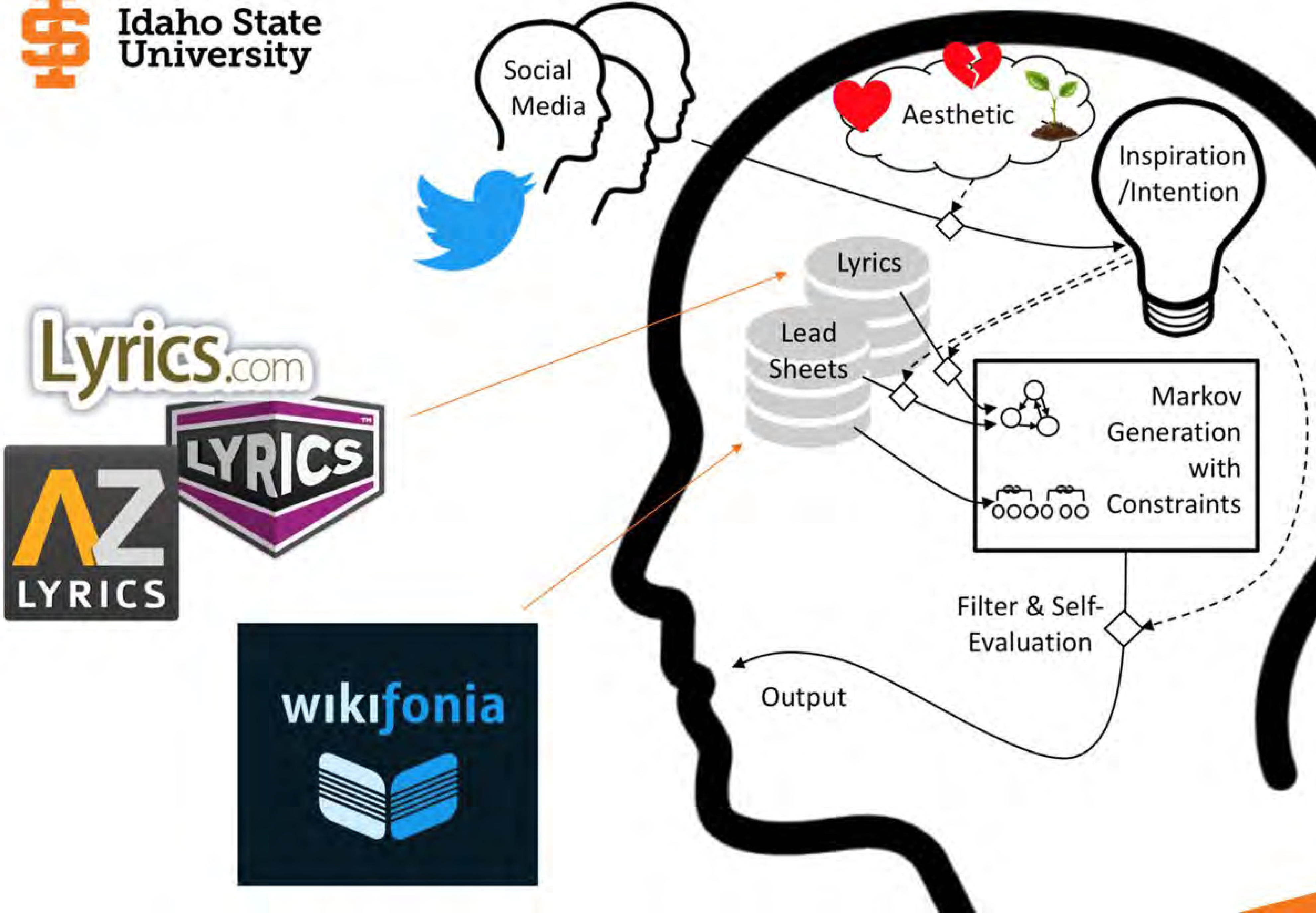
```
Prob:0.0034441160774746674, XProb:0.00357      Then you you you you can start to let It bet ter you have found her Then you you you you can start to
Prob:8.610290193686683E-4, XProb:9.4E-4 Then you you you you were made to let It bet ter you have found her Then you you you you were made to
Prob:0.01771259696986976, XProb:0.01837 let It bet ter you have found her Then you you you you can start to let It bet ter you have found her
Prob:0.02066469646484799 7, XProb:0.02055      Then you you you you can start to make It bet ter you have found her Then you you you you can start to
Prob:0.00401813542372045 5, XProb:0.00387      her Then you you you can start to go out and let and let and get her Then you you you can start to
Prob:3.587620914036119E-5, XProb:2.0E-5 ter you you you you can start to go out and let and let It bet ter you you you you can start to
Prob:0.5313779090960917, XProb:0.52992 make It bet ter you have found her Then you you you you can start to make It bet ter you have found her
Prob:0.13284447727402302, XProb:0.13195 make It bet ter you have found her Then you you you you begin to make It bet ter you have found her
Prob:0.0010045338559301102, XProb:8.7E-4       her Then you you you were made to go out and let and let and get her Then you you you were made to
Prob:7.175241828072212E-5, XProb:5.0E-5 ter you you you you were made to go out and let and make It bet ter you you you you were made to
Prob:0.00885629848493488, XProb:0.0087  let It bet ter you have found her Then you you you you were made to let It bet ter you have found her
Prob:0.005166174116212007, XProb:0.00497       Then you you you you were made to make It bet ter you have found her Then you you you you were made to
Prob:2.870096731228896E-4, XProb:2.1E-4 ter you you you you can start to go out and let and make It bet ter you you you you can start to
Prob:2.511334639825283 5E-4, XProb:3.0E-4      her Then you you you begin to go out and let and let and get her Then you you you begin to
Prob:0.0012915435290529994, XProb:0.00132      Then you you you you begin to make It bet ter you have found her Then you you you you begin to
Prob:3.690124368722867 7E-4, XProb:4.4E-4      go out and let and make It bet ter you you you you were made to go out and let and make It bet
Prob:7.380248737445736E-4, XProb:8.7E-4 go out and let and make It bet ter you you you you can start to go out and let and make It bet
Prob:1.845062184361433 8E-4, XProb:2.0E-4      go out and let and make It bet ter you you you you begin to go out and let and make It bet
Prob:0.2656889545480461, XProb:0.26728  make It bet ter you have found her Then you you you you were made to make It bet ter you have found her
Prob:2.152572548421666 6E-4, XProb:2.2E-4      Then you you you you begin to let It bet ter you have found her Then you you you you begin to
Prob:1.25566731991264 4E-4, XProb:1.5E-4 go out and let and let and get her Then you you you were made to go out and let and let and get
Prob:0.004428149242467439 5, XProb:0.00489     let It bet ter you have found her Then you you you you begin to let It bet ter you have found her
Prob:6.27833659956322E-5, XProb:1.0E-4  go out and let and let and get her Then you you you begin to go out and let and let and get
Prob:2.51133463982528 8E-4, XProb:2.4E-4 go out and let and let and get her Then you you you can start to go out and let and let and get
```

```
Prob:0.001190658097202680 4, XProb:0.00249     CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 CM7 Em7 Dm7 Bb7 CM7 Em7 Dm7 Bb7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7
Prob:1.708359299246263E-4, XProb:4.9E-4 CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 Em7 Dm7 Dm7 CM7 Em7 Dm7 Dm7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7
Prob:0.001860403276879188, XProb:0.00187       CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 Dm7 Bb7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7
Prob:9.187176675946564E-5, XProb:2.2E-4 CM7 CM7 Dm7 Bb7 CM7 CM7 CM7 CM7 CM7 Em7 Dm7 Dm7 CM7 Em7 Dm7 Dm7 CM7 CM7 Dm7 Bb7 CM7 CM7 CM7 CM7
Prob:9.187176675946597E-5, XProb:1.5E-4 CM7 Em7 Dm7 Dm7 CM7 CM7 CM7 CM7 CM7 CM7 Dm7 Bb7 CM7 CM7 Dm7 Bb7 CM7 Em7 Dm7 Dm7 CM7 CM7 CM7 CM7
Prob:0.004270898248115666 4, XProb:0.00891     CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 Em7 Dm7 G7 CM7 Em7 Dm7 G7 CM7 CM7 CM7 CM7 CM7 CM7 CM7 CM7
```

Imagine/Hey Jude

HARMONY: CM7 Am7 Em7 A7 Dm7 G7 CM7 CM7 CM7 Em7 Dm7 G7 CM7 Em7 Dm7 G7 CM7 Am7 Em7 A7 Dm7 G7 CM7 CM7
PITCH: G4 E4 G4 A3 A3 F4 F4 D4 B3 C5 C5 C4 C4 C5 C5 B4 G4 G4 G4 A4 A4 A4 A4 G4 G4 G4 G4 A4 A4 A4 A4 G4 G4 E4 G4 A3 …
RHYTHM: 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0 1.0 3.0
LYRICS: make It bet ter you have found her Then you you you you can start to make It bet ter you have found her

Pop*



Social Media
Aesthetic
Inspiration /Intention
Lyrics
Lead Sheets
Markov Generation with Constraints
Filter & Self-Evaluation
Output
Lyrics.com
LYRICS
AZ LYRICS
wikifonia

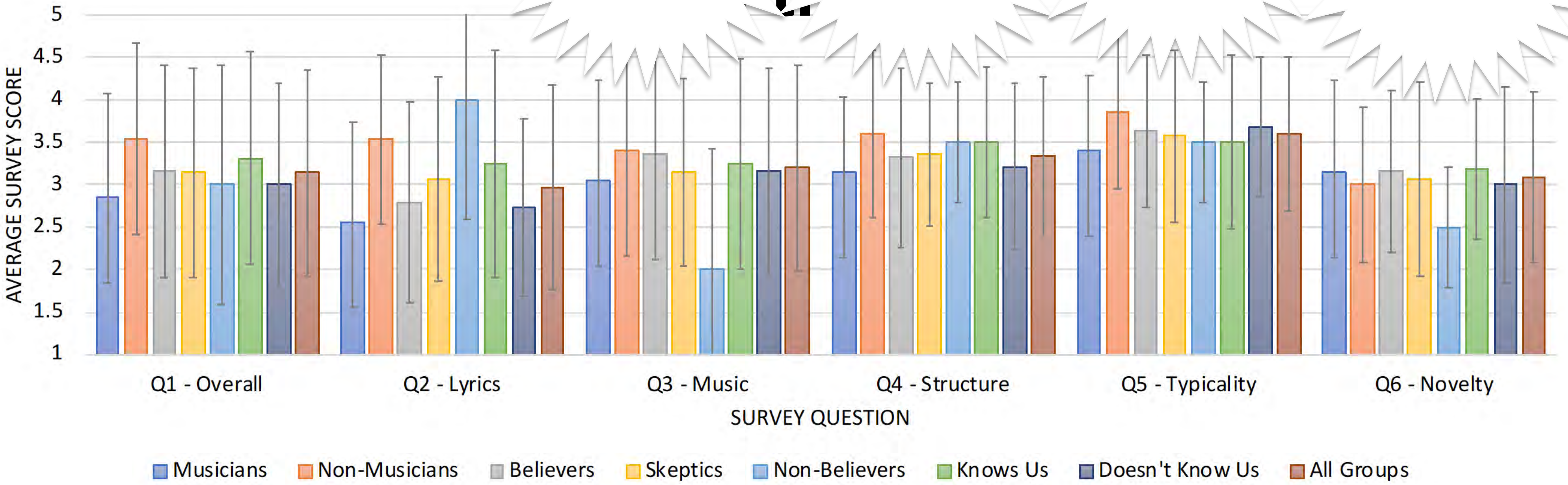
Challenge #4

Evaluating a creative AI system

Evaluation

For each of 12 randomly generated songs we asked:

1. How would you rate this **composition** overall?
2. How would you rate the **lyric** composition in this piece?
3. How would you rate the **music** composition in this piece?
4. How would you rate the **global structure** (i.e., form, layout) in this piece?
5. How **typical** is this song of pop/rock/show tunes music?
   Not at all typical ○ ○ ○ ○ ○ Very typical
6. How **novel** is this song compared to other pop/rock/show tunes music?
   I've heard this song before ○ ○ ○ ○ ○ I've never heard anything like this

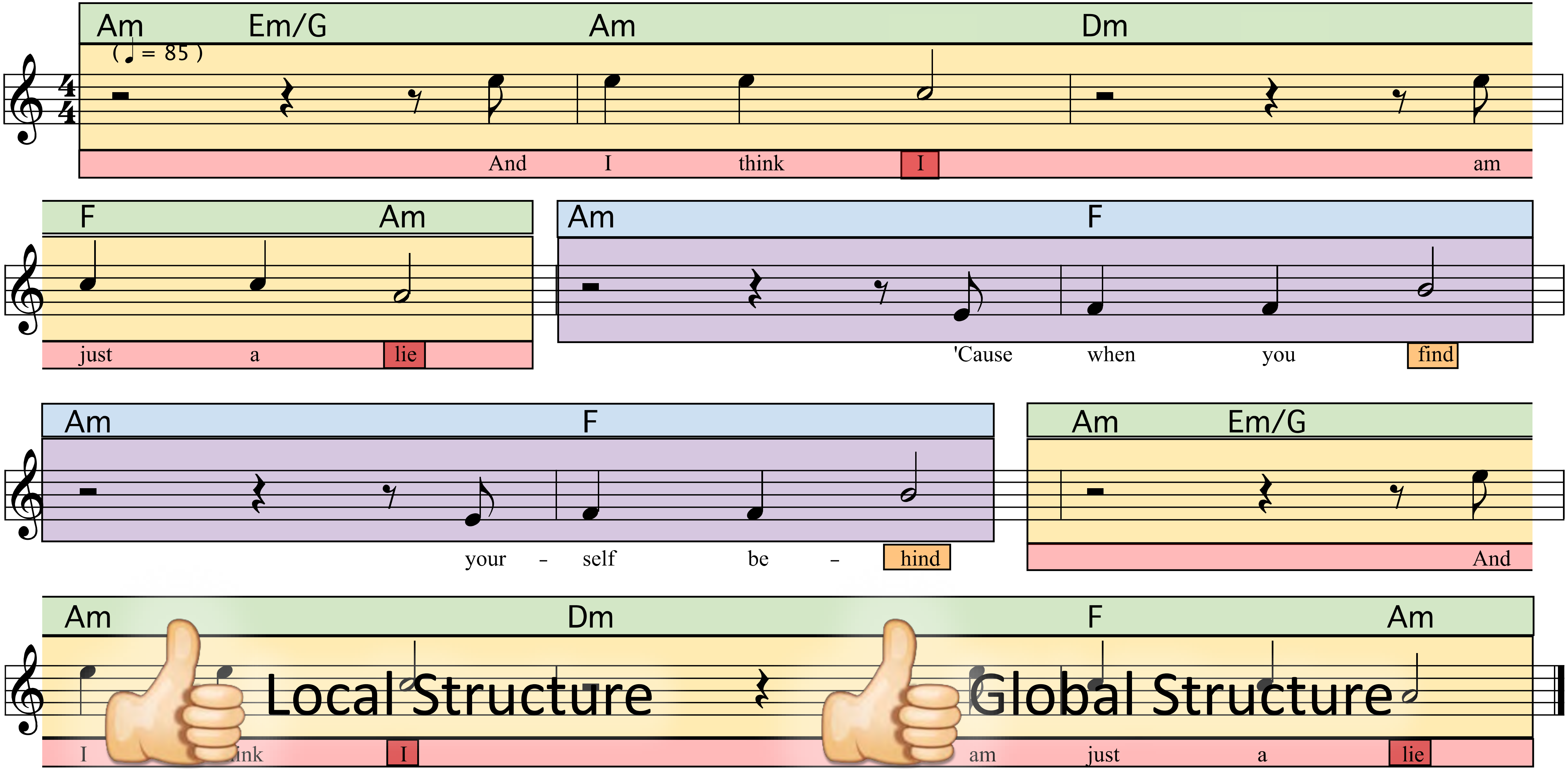


AVERAGE SURVEY SCORE
5
4.5
4
3.5
3
2.5
2
1.5
1
Q1 - Overall
Q2 - Lyrics
Q3 - Music
Q4 - Structure
Q5 - Typicality
Q6 - Novelty
SURVEY QUESTION
Musicians
Non-Musicians
Believers
Skeptics
Non-Believers
Knows Us
Doesn't Know Us
All Groups

And I think I

Inspired by: Fear and love

Pop*



Idaho State University

SONG CONTEST 2022

3rd International AI Song Contest, 2022

- A panel of international experts in music and AI evaluated **46 teams**
- The top 15 teams from this evaluation went on to a **public vote**
- The final score was determined by a **50–50 weighting** of the public vote and the second jury evaluation
- Pop* Team Collaborators:
  - **Dan Ventura** – BYU CS Professor and PhD Advisor
  - **Jon Armstrong** – ISU Music Professor
  - **Claire Smedley-Dye** – ISU Music student

ROAR


POLAR

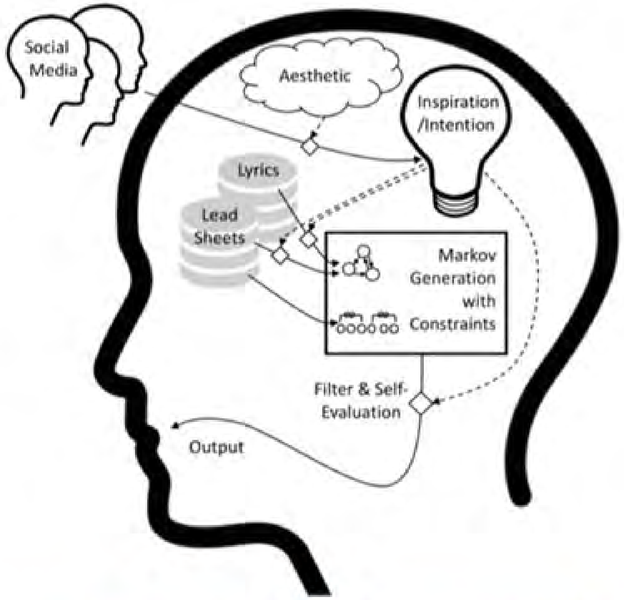

CANOPUS


CANOPUS


AND I THINK I

Idaho State University

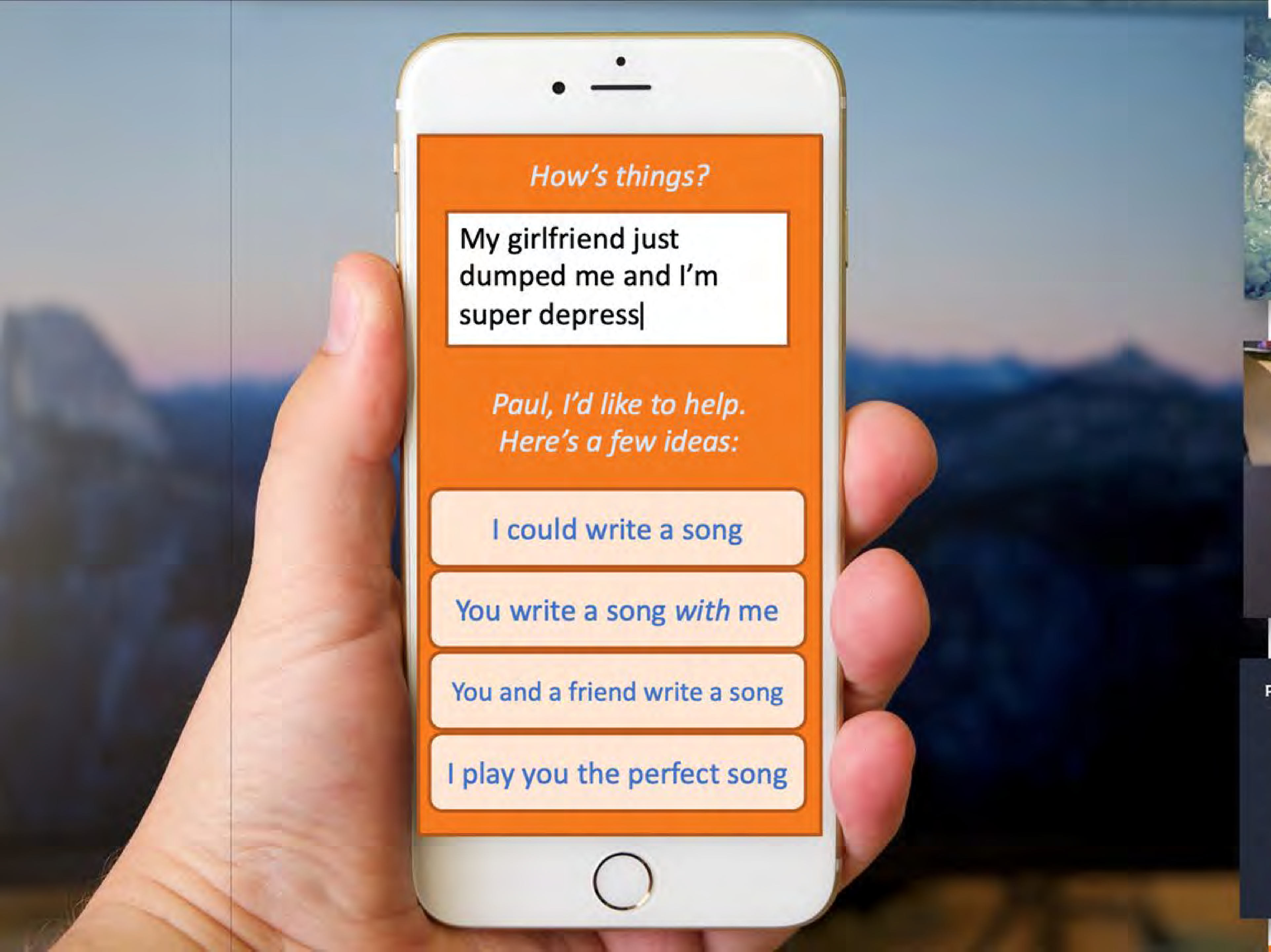
- Scores out of 12
- Pop*
  - **Jury**: 10th place
  - **Public**: 10th place

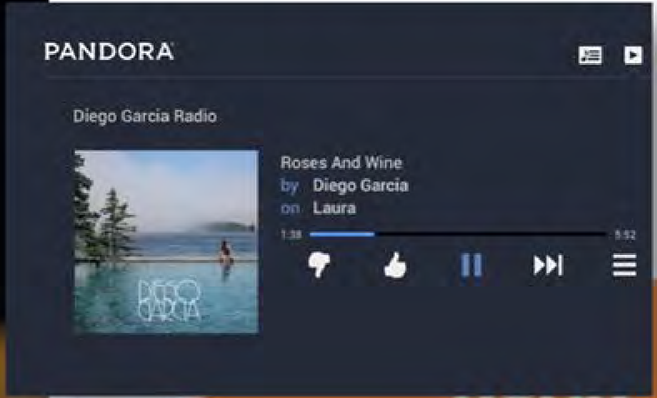


https://www.aisongcontest.com/winners-2022

Complete list of finalists

	Team Name	Total	Public	Jury
1	Yaboi Hanoi	21.1	10.3	10.8
2	PAMP!	20.4	11.3	9.1
3	3+i	19.6	11.2	8.4
4	Dadabots	19.2	9.8	9.4
5	BHC	18.8	10.6	8.3
6	Little Robots	18.7	10.5	8.2
7	AIM3	18.4	11.1	7.3
8	Wavy Weights and Bassy Biases	18.3	10.3	8.0
9	Rage against the Espresso Machine	17.9	8.8	9.2
10	Pop*	17.8	9.8	8.1
11	Machine Forgetting	17.6	9.7	7.9
12	G-Zone	17.2	10.1	7.1
13	Indy Rubbish Mice	16.0	9.2	6.8
14	Maskinspelmanslaget	14.0	8.3	5.7
15	LullabyBye	12.6	6	6.6


How's things?
My girlfriend just dumped me and I'm super depress
Paul, I'd like to help.
Here's a few ideas:
I could write a song
You write a song *with* me
You and a friend write a song
I play you the perfect song


PANDORA
Diego Garcia Radio
Roses And Wine
by Diego Garcia
on Laura

Thank you!

- Questions?

Paul Bodily
bodipaul@isu.edu

Accelerated Data Science and AI

Patty Delafuente | January 2023

Agenda

- RAPIDS
- NVIDIA AI Framework

CHALLENGES IN DATA SCIENCE TODAY

What Needs to be Solved to Empower Data Scientists

INCREASING BIG DATA USE CASES

Data sets are continuing to dramatically increase in size

Multitude of sources

Different data formats with varying quality

SLOW CPU PROCESSING

End of Moore's law, CPUs aren't getting faster

Many popular data science tools are CPU-only

CPUs are not designed to parallelize processes well

COMPLEX SOFTWARE MANAGEMENT

Time consuming to install software

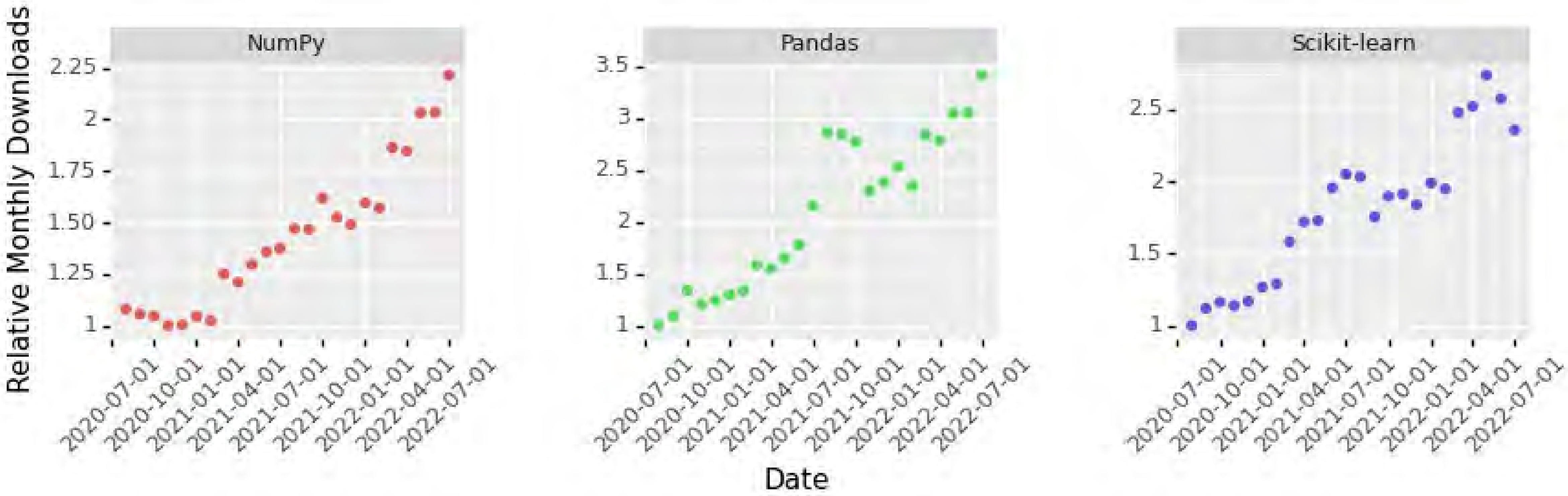
Nearly impossible to manage all version conflicts

Updates often break other software

ACCELERATING PYTHON DATA SCIENCE WITH RAPIDS

Continued Growth of Python Data Ecosystems

NumPy, Pandas, and Scikit-learn: 2x+ more monthly downloads than two years ago

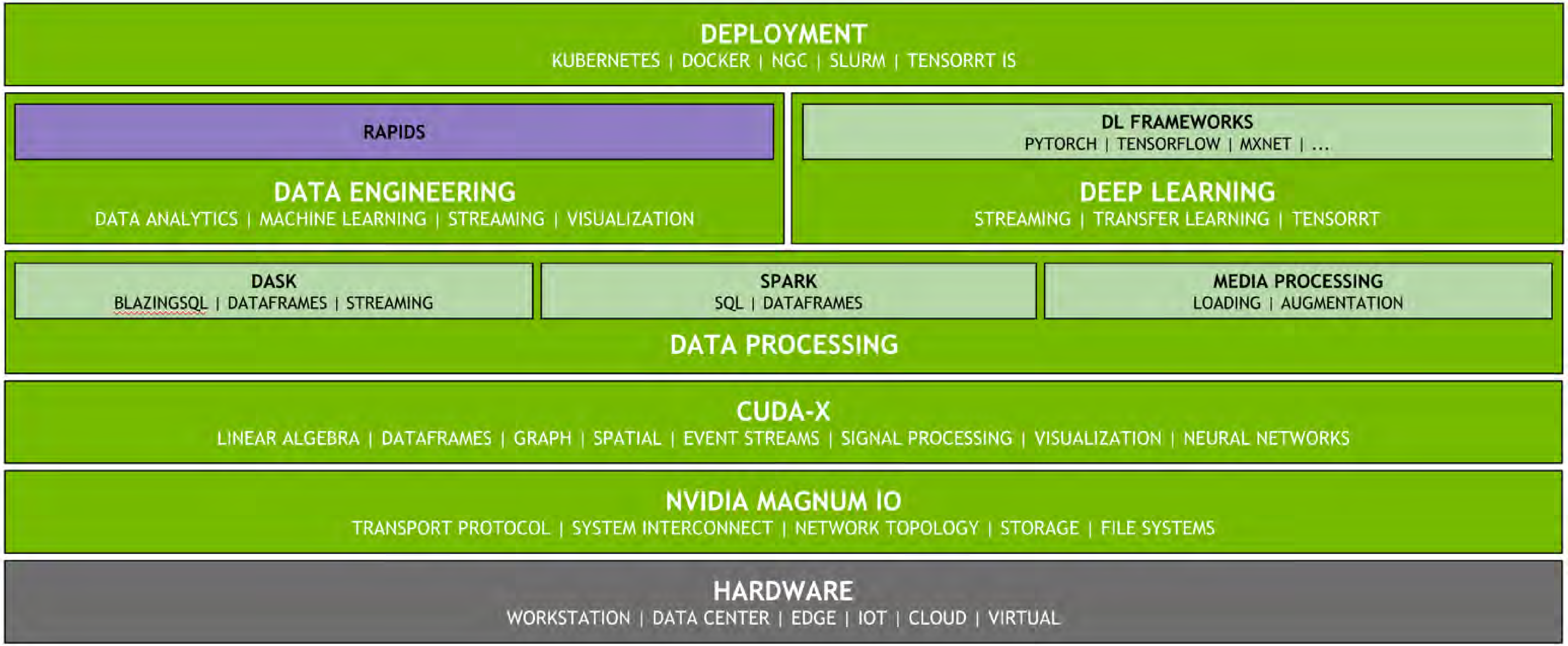


NVIDIA

"RAPIDS brings open-source GPU acceleration to data science and data engineering"

NVIDIA RAPIDS TRANSFORMS DATA SCIENCE

From Analytics to NVIDIA **Accelerated** Data Science

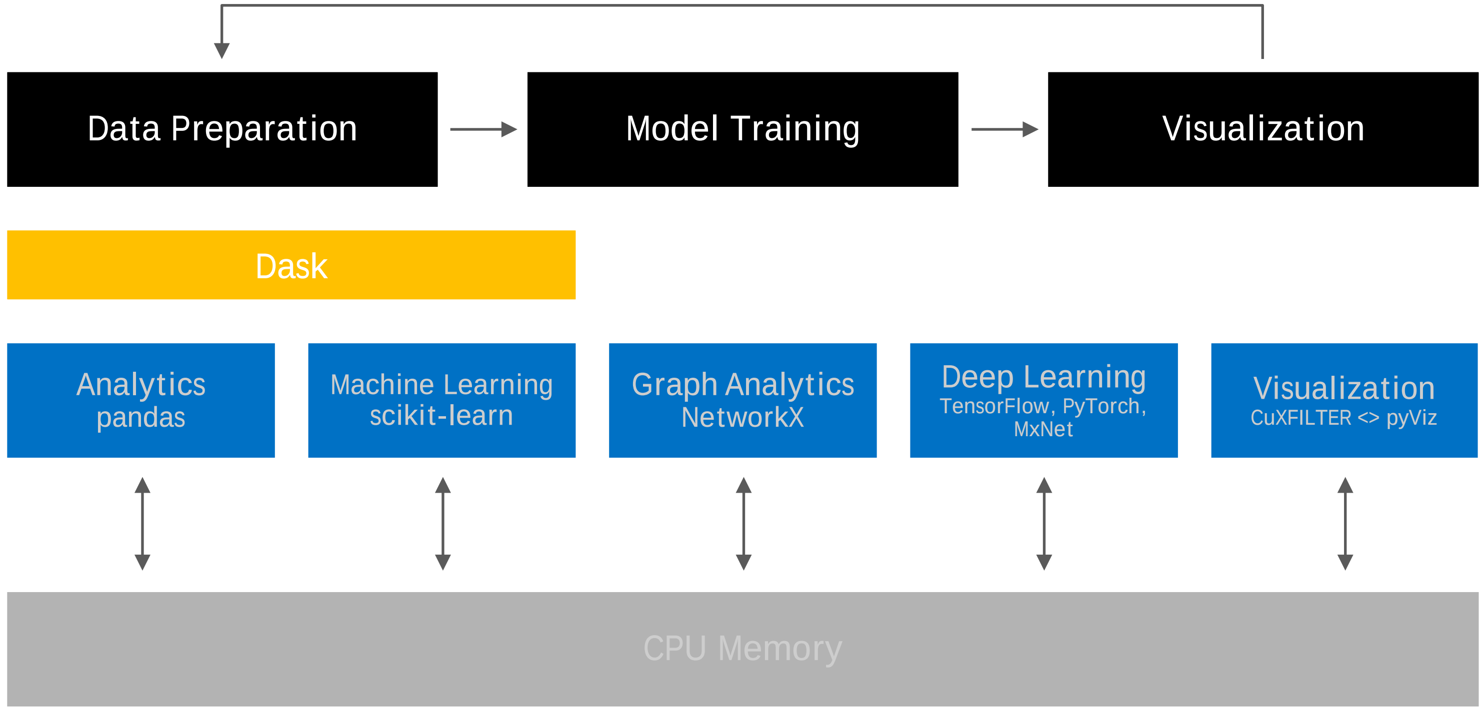


ACCESSIBLE, EASY TO USE TOOLS ABSTRACT COMPLEXITY

Python is the most-used language in Data Science today. Libraries like NumPy, Scikit-Learn, and Pandas have changed how we think about accessibility in Data Science and Machine Learning.

While great for experimentation, PyData tools lack the power necessary for enterprise-scale workloads. This leads to substantial refactoring to handle the size of modern problems, increasing cycle time, overhead, and time to insight.

These pain points are further compounded by computational bottlenecks of CPU-based processing.



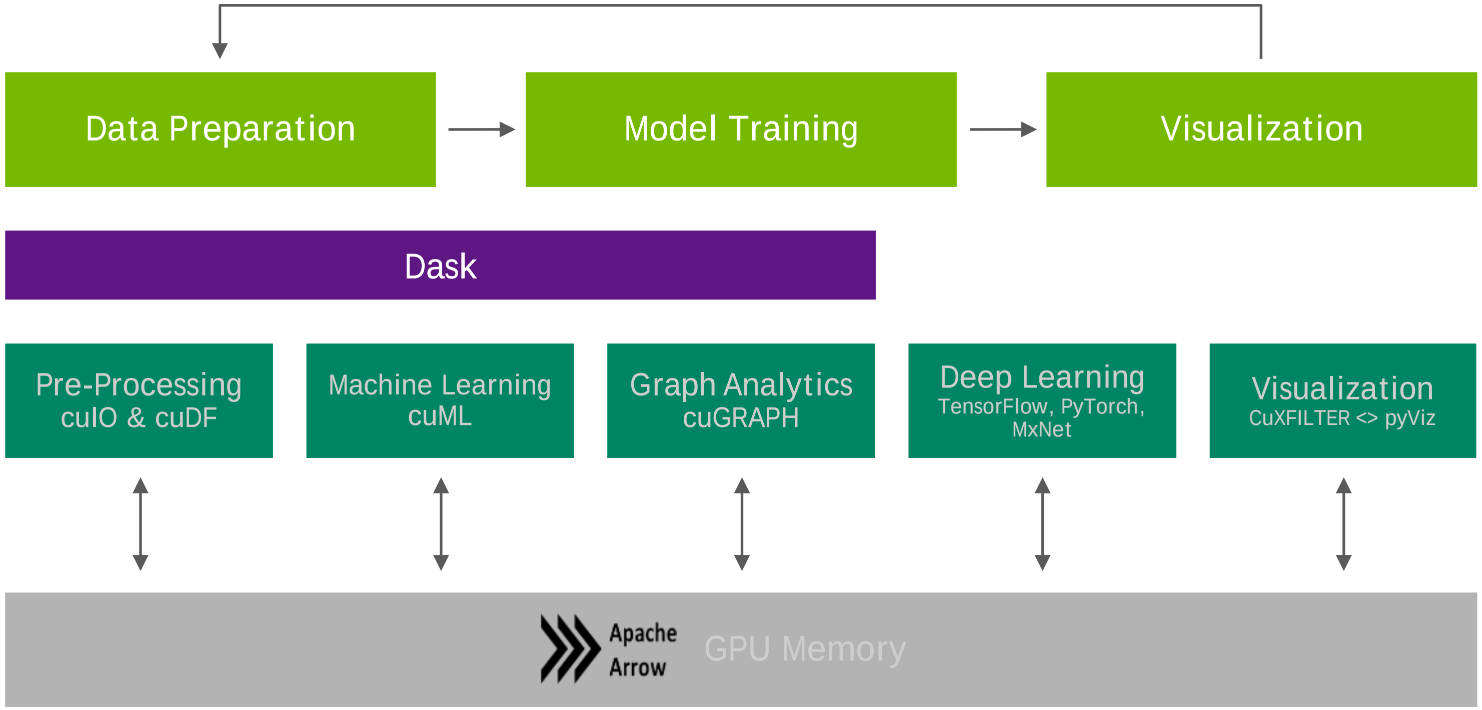
Code refactors and inter-team handoffs decrease data-driven ROI

DELIVERING ENTERPRISE-GRADE DATA SCIENCE SOLUTIONS IN PURE PYTHON

The RAPIDS suite of open source software libraries gives you the freedom to execute end-to-end data science and analytics pipelines entirely on GPUs.

RAPIDS utilizes **NVIDIA CUDA** primitives for low-level compute optimization and exposes GPU parallelism and high-bandwidth memory speed through user-friendly Python interfaces like PyData.

With Dask, RAPIDS can scale out to multi-node, multi-GPU cluster to power through big data processes.



RAPIDS enables the PyData stack with the power of NVIDIA GPUs

DISTRIBUTE & ACCELERATE COMPUTATION FOR PRODUCTION WORKLOADS

RAPIDS

Accelerates PyData on NVIDIA GPUs

NumPy -> CuPy/PyTorch/..
Pandas -> cuDF
Scikit-Learn -> cuML
Numba -> Numba

RAPIDS + DASK

Distributes and accelerates PyData

Can be distributed across Multi-GPU on single node (DGX) or across a cluster

Provides easy to use tooling enabling HPC-level performance

PYDATA

Provides accessible, easy to use tooling

NumPy, Pandas, Scikit-Learn, Numba and many more

Single CPU core, in-memory data

DASK

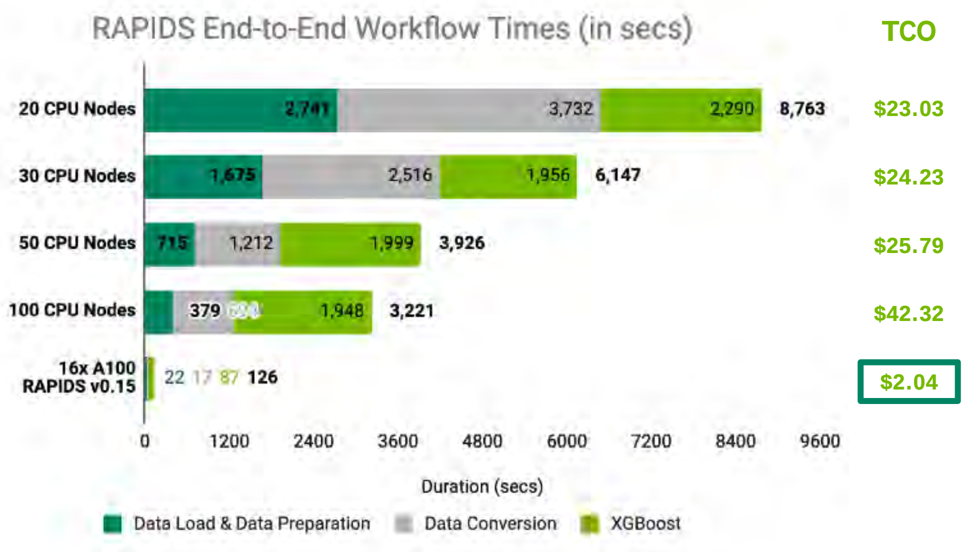
Distributes PyData across multiple cores

NumPy -> Dask Array
Pandas -> Dask DataFrame
Scikit-Learn -> Dask-ML
… -> Dask Futures

REDUCING DATA SCIENCE PROCESSES FROM HOURS TO SECONDS

RAPIDS delivers massive speed-ups across the end-to-end data science lifecycle. Conducting benchmarks in a commercial cloud environment, we're able to get incredible performance running a common ML model training pipeline.

Between loading and cleansing data, engineering features, and training a classifier using a 200GB CSV dataset, a RAPIDS-based pipeline completed these operations in ***just over two minutes***. The same process takes two and half hours on a similar CPU-configuration.



16

A100s Provide More Power than 100 CPU Nodes*

70x

Faster Performance than Similar CPU Configuration

20x

More Cost-Effective than Similar CPU Configuration

*CPU approximate to n1-highmem-8 (8 vCPUs, 52GB memory) on Google Cloud Platform.

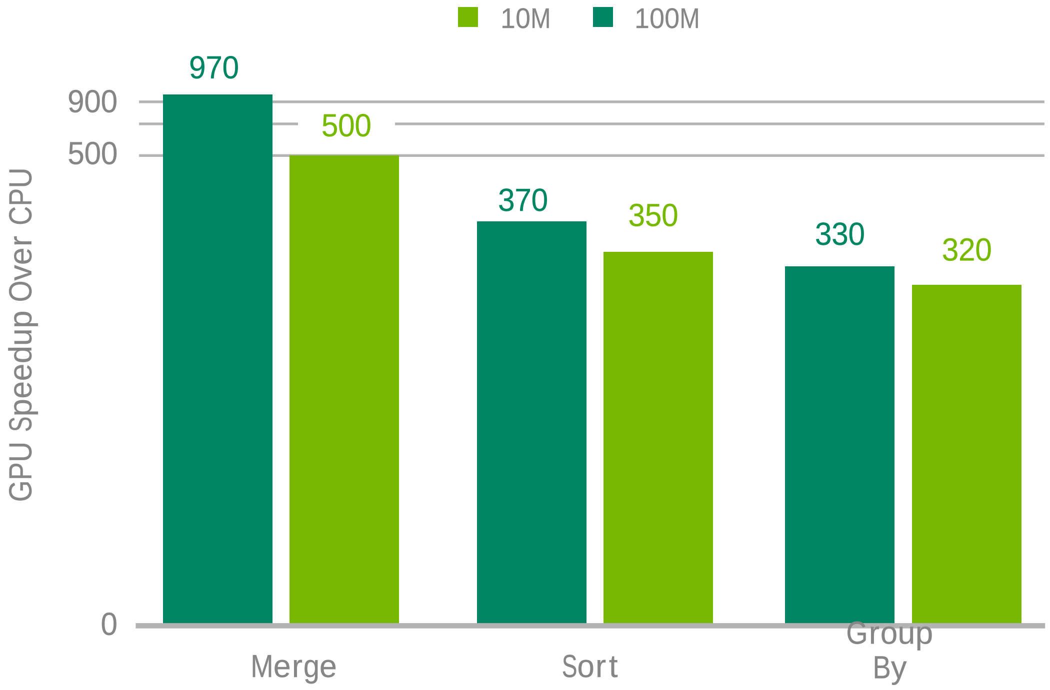
NVIDIA.

ACCELERATED PRE-PROCESSING

A FAMILIAR EXPERIENCE FOR PYTHON AND PANDAS USERS

cuDF is a Python GPU DataFrame library for loading, joining, aggregating, filtering, and otherwise manipulating data all in a **pandas-like** API familiar to data scientists.

With cuDF, users can create GPU DataFrames from Numpy arrays, Pandas DataFrames, and PyArrow Tables. Once in cuDF format, users can utilize other GPU-accelerated libraries to easily conduct machine learning and analytics processes.



Single GPU Speed-Ups vs Pandas

GPU: NVIDIA Tesla V100 32GB on DGX-1
CPU: Intel(R) Xeon(R) CPU E5-2698 v4 @ 2.20GHz

ACCELERATED MACHINE LEARNING

GPU-POWER WITH THE FEEL OF SCIKIT-LEARN

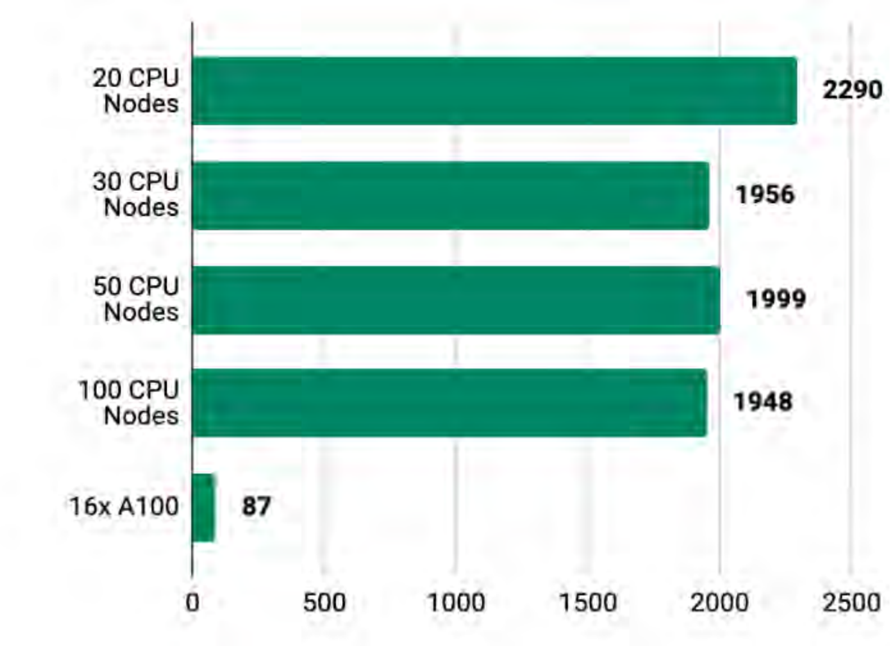
cuML is a suite of libraries that implement machine learning algorithms and mathematical primitives functions that share compatible APIs with other RAPIDS projects.

cuML enables data scientists, researchers, and software engineers to run traditional tabular ML tasks on GPUs without going into the details of CUDA programming. In most cases, cuML's Python API matches the API from scikit-learn.

26 GPU-Accelerated Algorithms & Growing

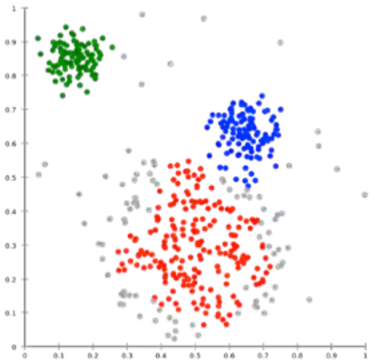
XGBoost Training Improvements

GPU: NVIDIA Ampere DGX A100s
CPU: 61GB memory, 8 vCPUs, 64-bit platform



Algorithms

GPU-accelerated Scikit-Learn



Cross Validation

Hyper-parameter Tuning

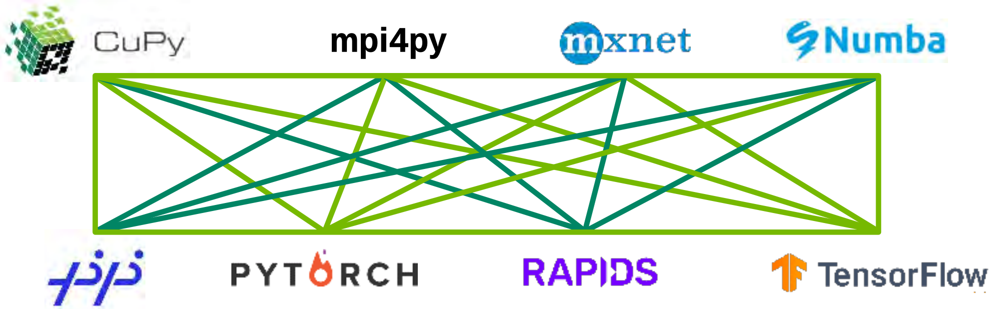
More to come!

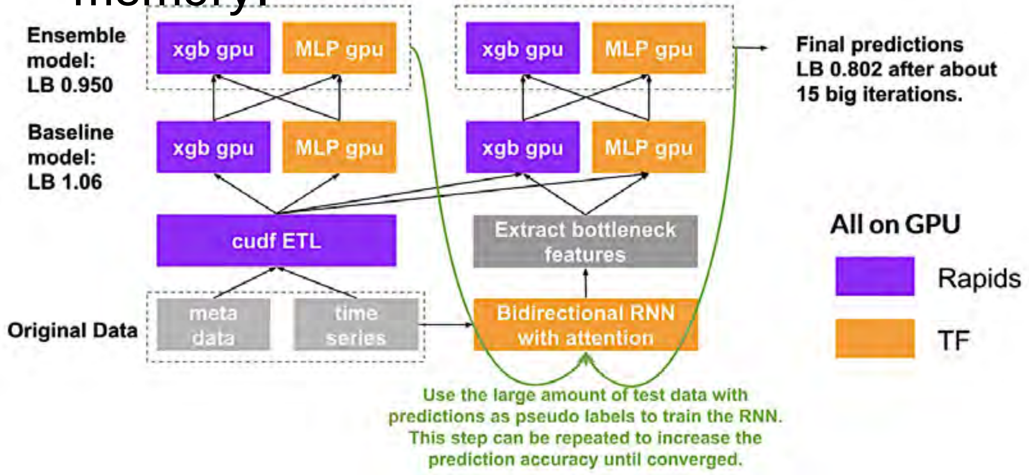
Classification / Regression	Decision Trees / Random Forests Linear/Lasso/Ridge/ElasticNet Regression Logistic Regression K-Nearest Neighbors Support Vector Machine Classification and Regression Naive Bayes
Inference	Random Forest / GBDT Inference (FIL)
Preprocessing	Text vectorization (TF-IDF / Count) Target Encoding Cross-validation / splitting
Clustering Decomposition & Dimensionality Reduction	K-Means DBSCAN Spectral Clustering Principal Components Singular Value Decomposition UMAP Spectral Embedding T-SNE
Time Series	Holt-Winters Seasonal ARIMA / Auto ARIMA

https://github.com/rapidsai/cuml#supported-algorithms

BUILD COMPLEX WORKFLOWS WITHOUT LEAVING THE GPU

RAPIDS supports device memory sharing between many popular data science and deep learning libraries, such as PyTorch and TensorFlow. By providing native array interface support, data can stay on the GPU avoiding costly copying back and forth to host memory.





Data stored in Apache Arrow can be seamlessly pushed to deep learning frameworks that accept CUDA Array Interface protocol or work with DLPack, such as Chainer, MXNet, and PyTorch.

NVIDIA Morpheus

Prebuilt Models for Data Driven Cybersecurity

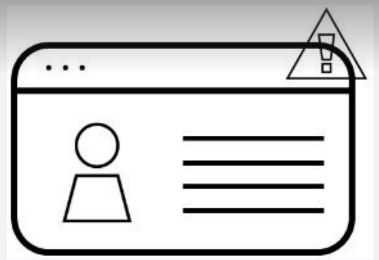
DIGITAL FINGERPRINTING
Analyze Behavior of Every User and Machine

PHISHING DETECTION
Analyze Raw Emails to Classify Automatically

LEAKED SENSITIVE INFORMATION
Find and classify leaked credentials

ANOMALOUS BEHAVIOR PROFILING
Profile behaviors to detect malicious code

IDENTITY FRAUD
Leveraging Graph Neural Networks



NVIDIA cuOPT

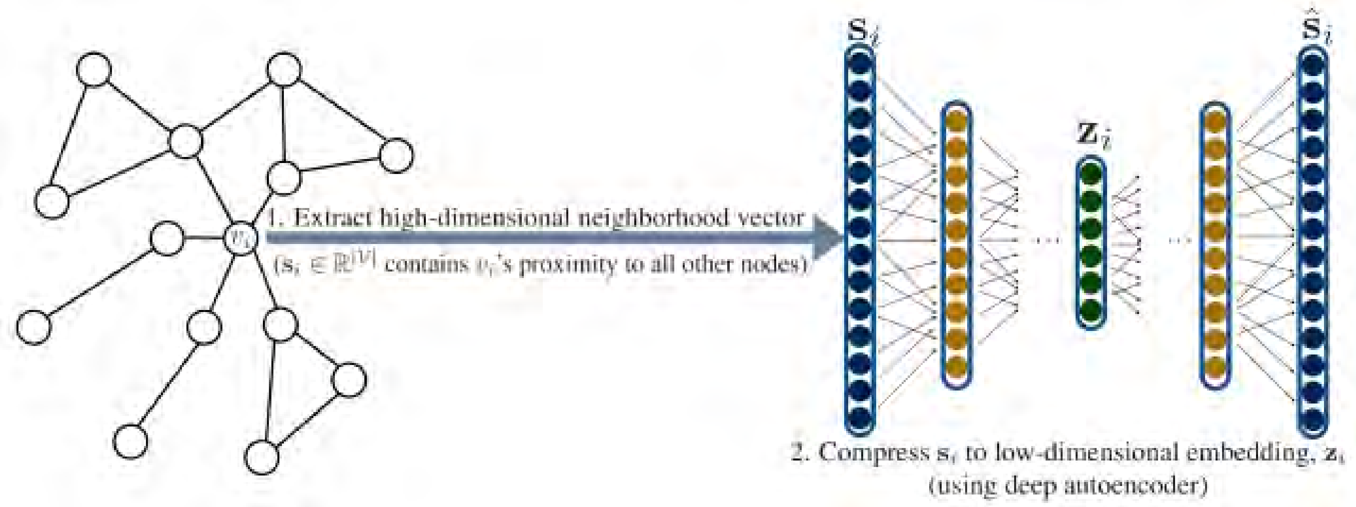
Fast, Accurate, and Scalable Optimization

- NVIDIA cuOpt™ is a GPU-accelerated solver for complex vehicle routing problems with a wide range of constraints
- Pick up and deliveries : goods need to be moved from certain pickup locations to other delivery locations. Find optimal visit sequence for a fleet of associates to visit the pickup and drop-off locations.
- Patented parallel heuristics on GPU with parallel compute
- Competitive advantage : speed and accuracy to enable dynamic re-optimization at scale

Graph Neural Networks

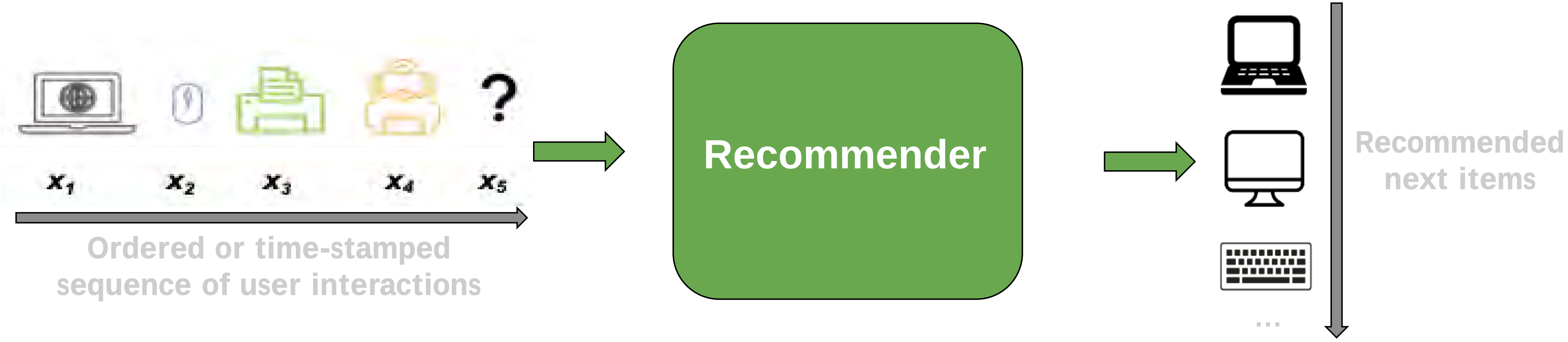
RAPIDS + Deep Learning = Performance and Scalability

- New containers on NGC provide the latest GPU acceleration for DGL and PyTorch Geometric
- Single A100 up to 9.5x faster than 128 CPU cores for RGCN model on MAG240 (vs. 2x AMD7268 EPYC)
- Mix and match GNNs and traditional ETL and graph algorithms smoothly with full cuGraph and cuDF integration



Transformers4Rec

A toolkit for sequential and session based recommendation



- **Sequential recommendation** - Leverages user past interactions (usually long sequences)
- **Session-based recommendation** - Leverages only user interactions within current session (usually short sequences)

RAPIDS Accelerator for Apache Spark

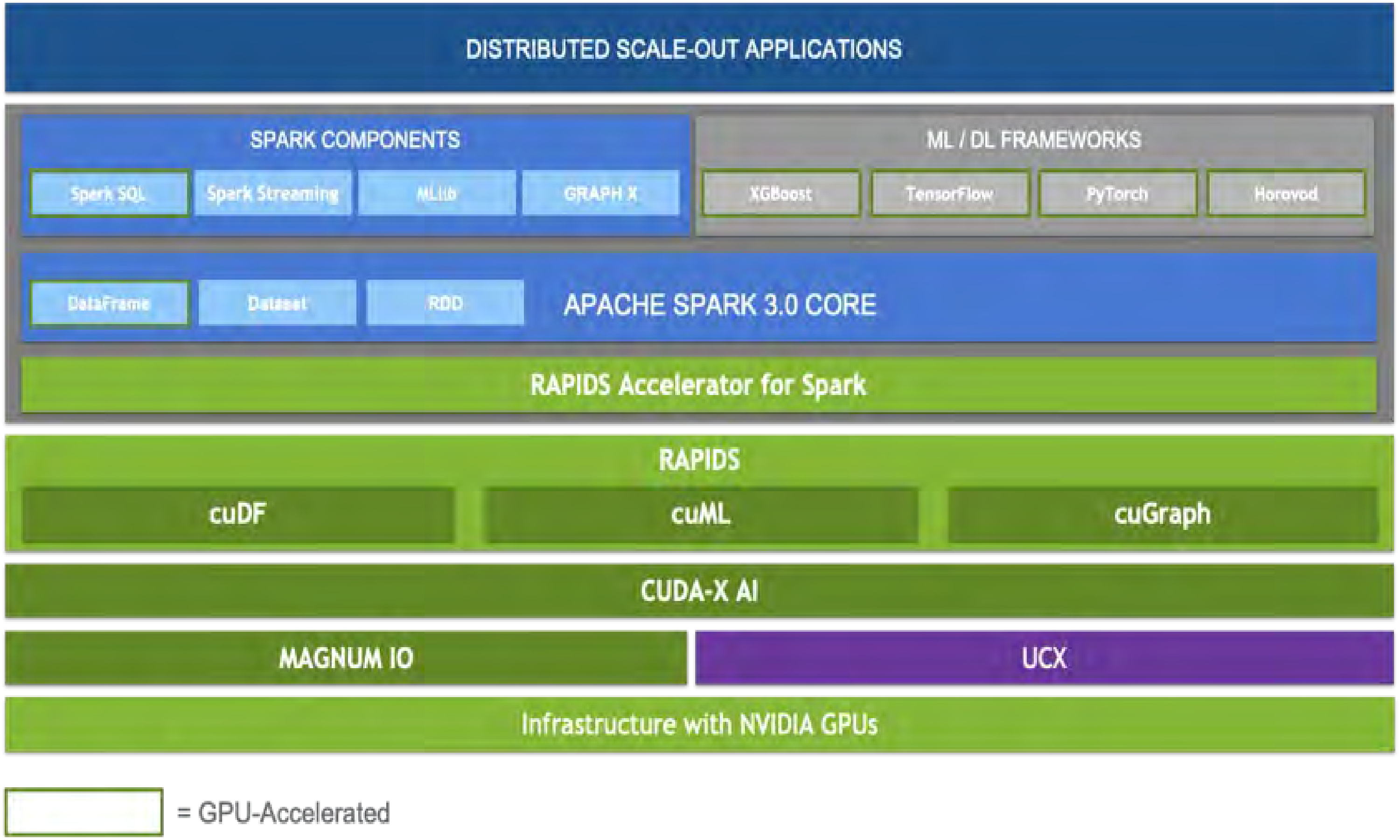
Seamless integration with Apache Spark 3.x

Features

- Use existing (unmodified) customer code
- Spark features that are not GPU enabled run transparently on the CPU

GPU Acceleration of:

- Spark Data Frames
- Spark SQL
- Spark ML



NO QUERY CHANGES

- Add jars to classpath and set spark plugins config
- Same SQL and DataFrame code
- Compatible with PySpark, SparkR, Koalas, and other DataFrame-based APIs
- Seamless fallback to CPU for unsupported operations

```
spark.sql("""
SELECT
  o_order_priority
  count(*) as order_count
FROM
  orders
WHERE
  o_orderdate >= DATE '1993-07-01'
  AND o_orderdate < DATE '1993-07-01' + interval '3' month
  AND EXISTS (
    SELECT
      *
    FROM
      lineitem
    WHERE
      l_orderkey = o_orderkey
      AND l_commitdate < l_receiptdate
  )
GROUP BY
  o_orderpriority
ORDER BY
  o_orderpriority
""").show()
```

HOW TO GET STARTED WITH RAPIDS

A VARIETY OF WAYS TO GET UP & RUNNING

More about RAPIDS

- Learn more at RAPIDS.ai
- Read the API docs
- Check out the RAPIDS blog
- Read the NVIDIA DevBlog

Self-Start Resources

- Get started with RAPIDS
- Deploy on the Cloud today
- Start with Google Colab
- Look at the cheat sheets

Discussion & Support

- Check the RAPIDS GitHub
- Use the NVIDIA Forums
- Reach out on https://rapids.ai/slack-invite
- NVIDIA Services

AI Frameworks and Tools

NVIDIA End-to-End AI Software Suite

Deep Learning Streamlined From Conception to Production at Scale

DATA PREP

NVIDIA RAPIDS

- Reduces data science processes from hours to seconds
- 70x faster performance than similar CPU configuration
- 20x more cost-effective than similar CPU configuration

TRAIN AT SCALE

NVIDIA TAO TOOLKIT

PyTorch

TensorFlow

- Train, Adapt, Optimize Models in hours vs. months
- Open-source ML frameworks optimized for GPU
- Integrated w/ NVIDIA RAPIDS to simplify development

OPTIMIZED FOR INFERENCE

TensorRT

- Maximize throughput for latency-critical apps w/ compiler & runtime
- Optimize every network (CNNs, RNNs, & Transformers)
- Optimizes use of GPU memory bandwidth

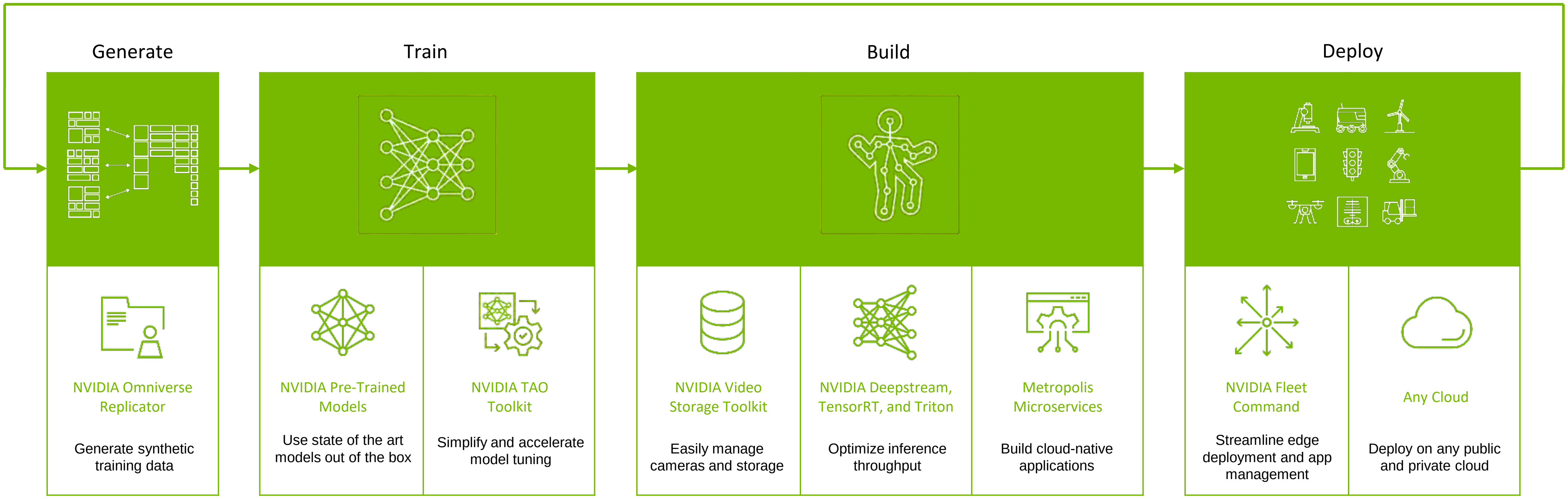
DEPLOY AT SCALE

NVIDIA TRITON INFERENCE SERVER

- Fast & scalable AI to applications
- Diverse query types – real time, offline batch, ensembles
- Up to 266x performance increase over CPU-only
- Triton with FIL backend delivers best inference performance for tree-based models on GPUs

End-to-End Vision AI Development

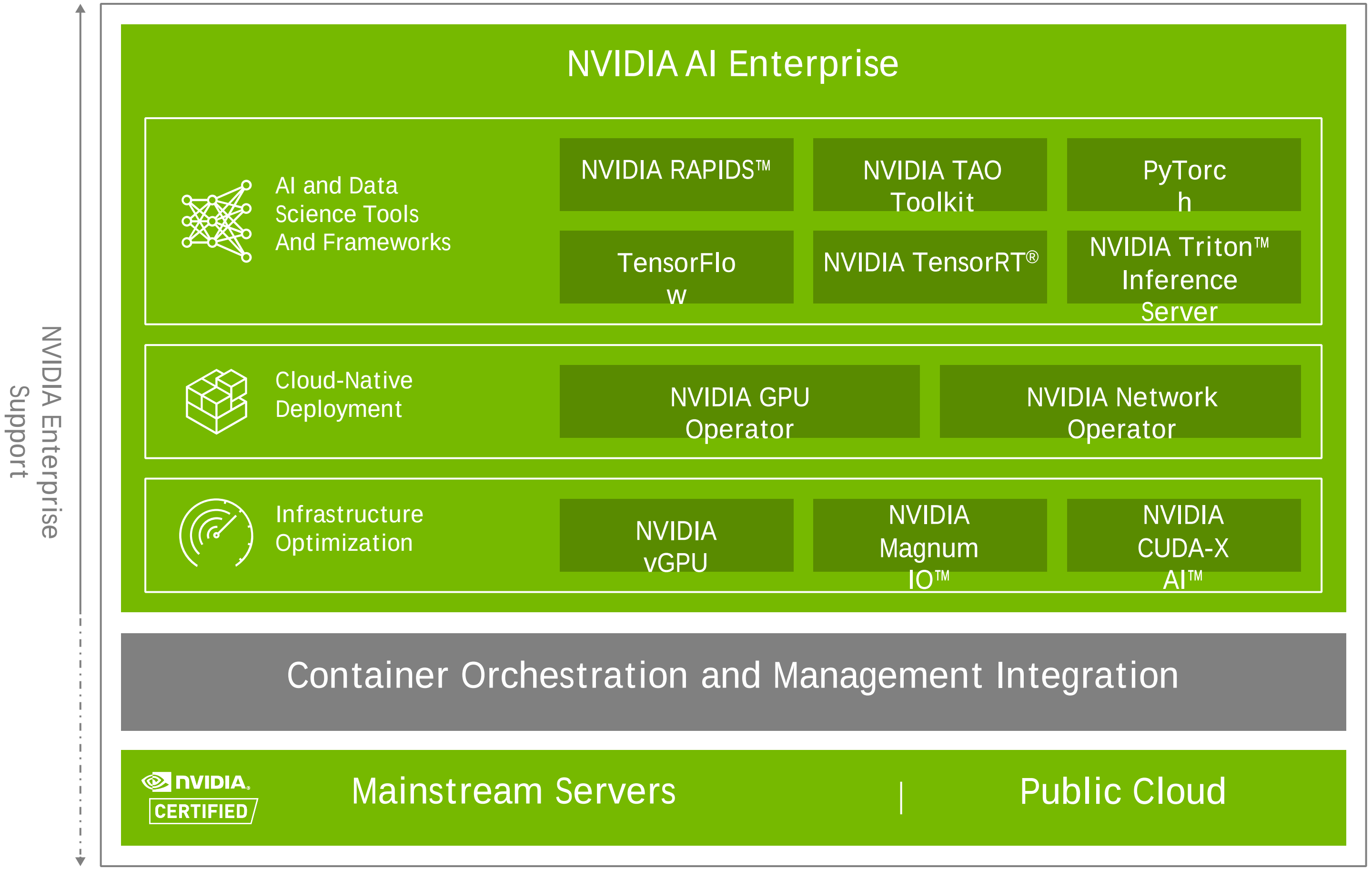
Fast-Track Data Generation, AI Model Creation, App Development, Inference and Scalability



NVIDIA

NVIDIA AI ENTERPRISE SOFTWARE SUITE

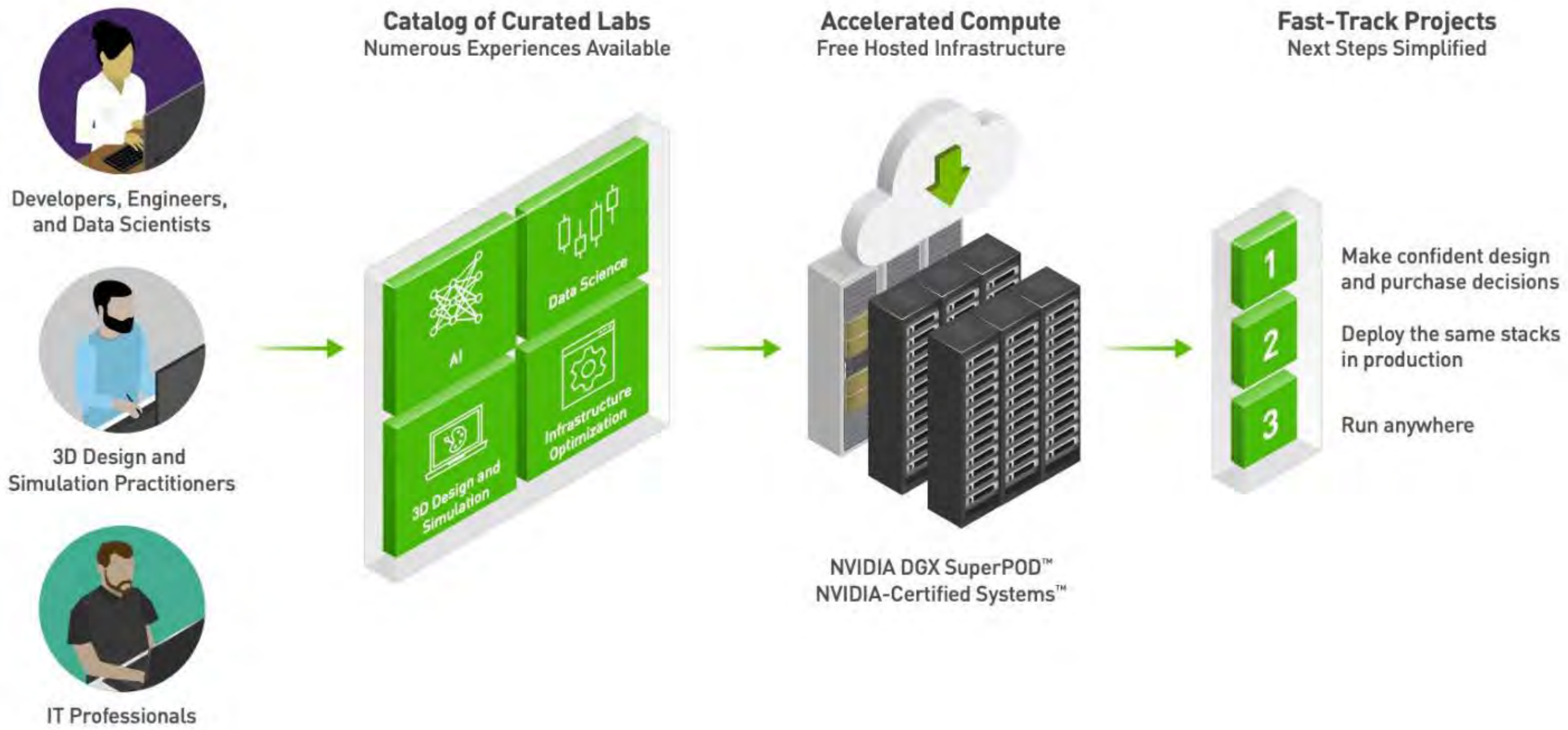
Enterprise AI for Everyone, Everywhere, on Every Platform



https://www.nvidia.com/ai-enterprise-suite/

NVIDIA LaunchPad

Instantly experience end-to-end workflows for AI, data science, 3D design collaboration, and more



Get Started at nvidia.com/launchpad

NVIDIA

AMD INSTINCT™ & AI

Ian Ferreira

Confidential & Under NDA

[Confidential - Distribution with NDA]

CAUTIONARY STATEMENT

This presentation contains forward-looking statements concerning Advanced Micro Devices, Inc. (AMD) such as the features, functionality, performance, availability, timing and expected benefits of AMD products; and the AMD Instinct™ roadmap, which are made pursuant to the Safe Harbor provisions of the Private Securities Litigation Reform Act of 1995. Forward-looking statements are commonly identified by words such as "would," "may," "expects," "believes," "plans," "intends," "projects" and other terms with similar meaning. Investors are cautioned that the forward-looking statements in this presentation are based on current beliefs, assumptions and expectations, speak only as of the date of this presentation and involve risks and uncertainties that could cause actual results to differ materially from current expectations. Such statements are subject to certain known and unknown risks and uncertainties, many of which are difficult to predict and generally beyond AMD's control, that could cause actual results and other future events to differ materially from those expressed in, or implied or projected by, the forward-looking information and statements. Investors are urged to review in detail the risks and uncertainties in AMD’s Securities and Exchange Commission filings, including but not limited to AMD’s most recent reports on Forms 10-K and 10-Q.

AMD does not assume, and hereby disclaims, any obligation to update forward-looking statements made in this presentation, except as may be required by law.

[Confidential - Distribution with NDA]

MISSION

Together we advance AI to help solve the world's most important challenges

PERFORMANT

Competitive performance helps ensure we meet the growing needs of accelerated computing

INCLUSIVE

An open software stack and competitive pricing maximizes reach and participation

RESPONSIBLE

Power efficiency helps reduce the environmental impact of accelerated computing

[Confidential - Distribution with NDA]

OUR HERITAGE IN HIGH PERFORMANCE COMPUTING

TOP500 The List.

#1 FRONTIER

#3 LUMI

#10 GENCI CINES

THE GREEN500

#1 Oak Ridge National Laboratory

#2 FRONTIER

#3 LUMI

#4 GENCI CINES

Oak Ridge National Laboratory · Lawrence Livermore National Laboratory · Vetenskap och Konst Kungl Tekniska Högskolan · PAWSEY supercomputing centre · GENCI · Nikhef · CGG · eni · NCSA

Top500 and Green500 lists as of June 2022

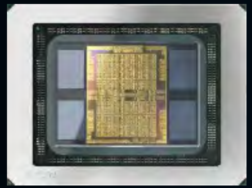
HARDWARE

Our Leadership in Silicon

together we advance_

[Confidential - Distribution with NDA]

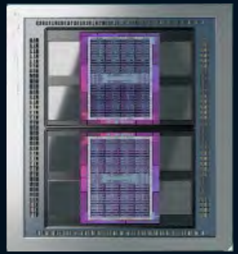
AMD INSTINCT™ JOURNEY ON ACCELERATED COMPUTING



AMD Instinct™ MI100
AMD CDNA™

Ecosystem Growth

First purpose-built GPU architecture for the data center



AMD Instinct™ MI200
AMD CDNA™ 2

Driving HPC and AI to a New Frontier

First purpose-built GPU powering discovery at Exascale

AMD Instinct™ MI300
AMD CDNA™ 3

Data Center APU

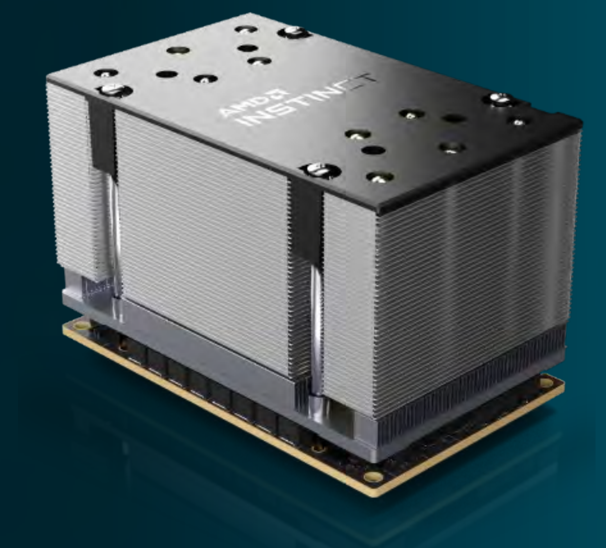
Breakthrough architecture designed for leadership efficiency and performance for HPC and AI

2020 → 2023

Roadmaps Subject to Change

[Confidential - Distribution with NDA]

AMD INSTINCT™ 200 FAMILY



AMD INSTINCT™
MI200 OAM SERIES

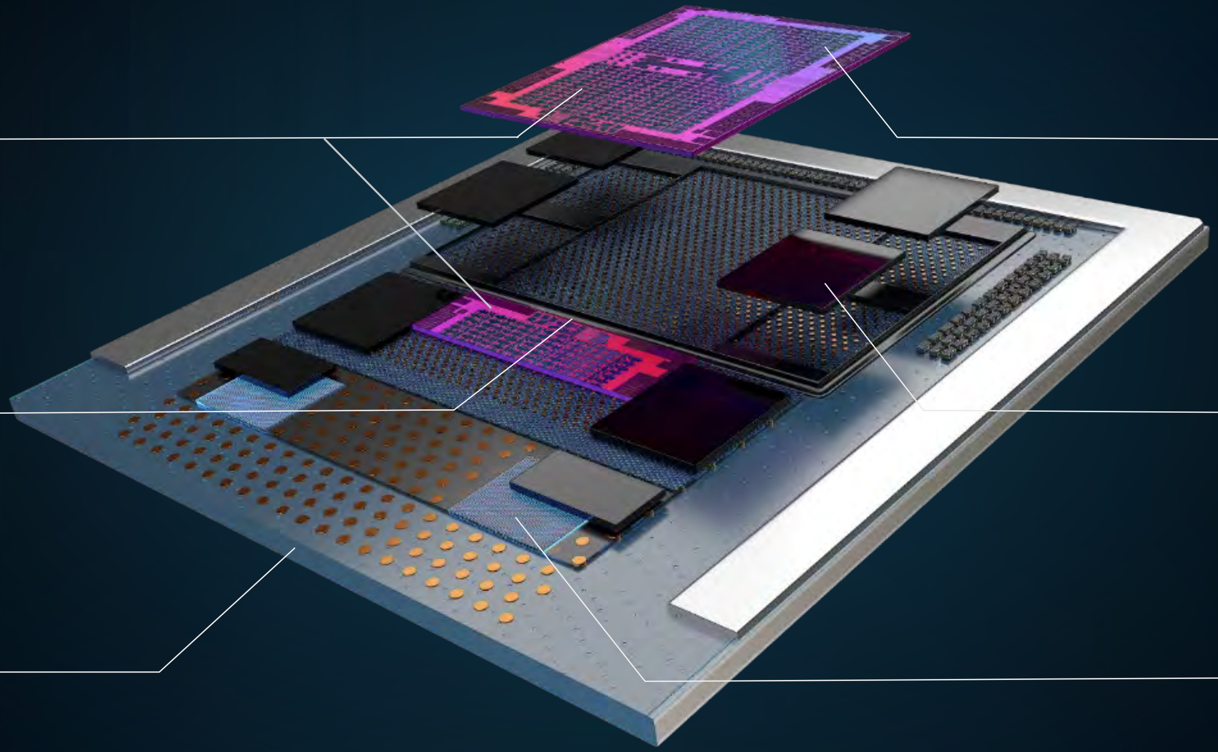
MI250, MI250X

AMD INSTINCT™
MI200 PCIe® SERIES

MI210

[Confidential - Distribution with NDA]

AMD INSTINCT™ MI-250 INNOVATIONS



AMD INSTINCT™ MI200 OAM SERIES

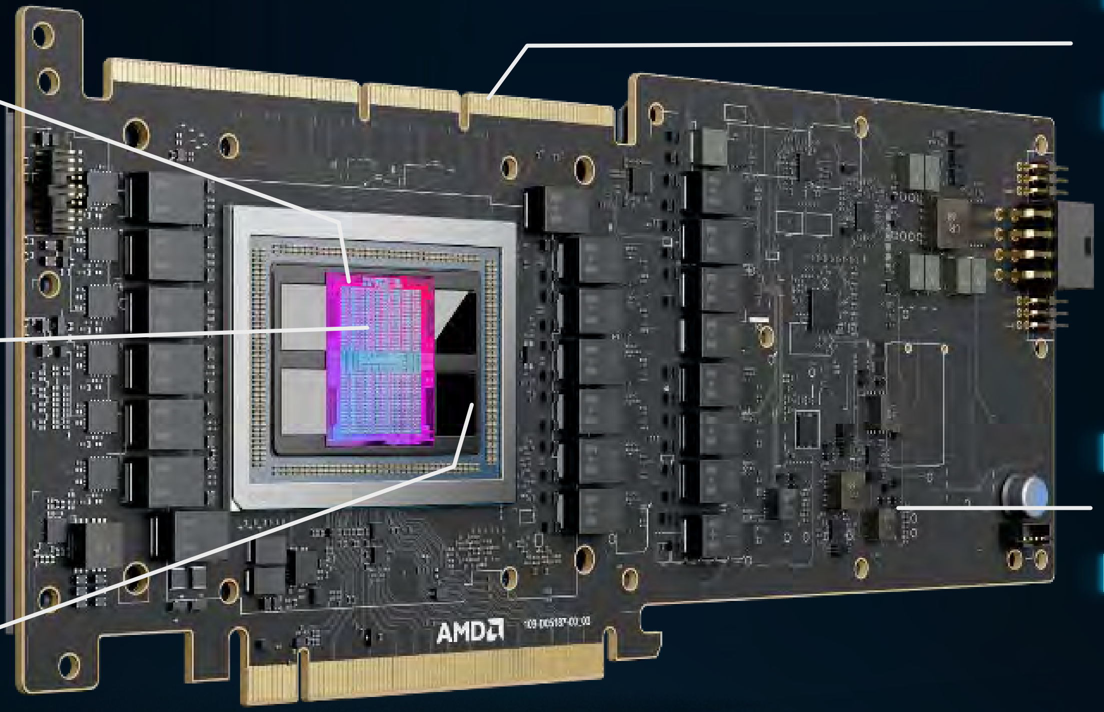
[Confidential - Distribution with NDA]

AMD INSTINCT™ MI-210 INNOVATIONS

AMD CDNA™ 2
GRAPHICS COMPUTE DIE

2nd GEN MATRIX CORES

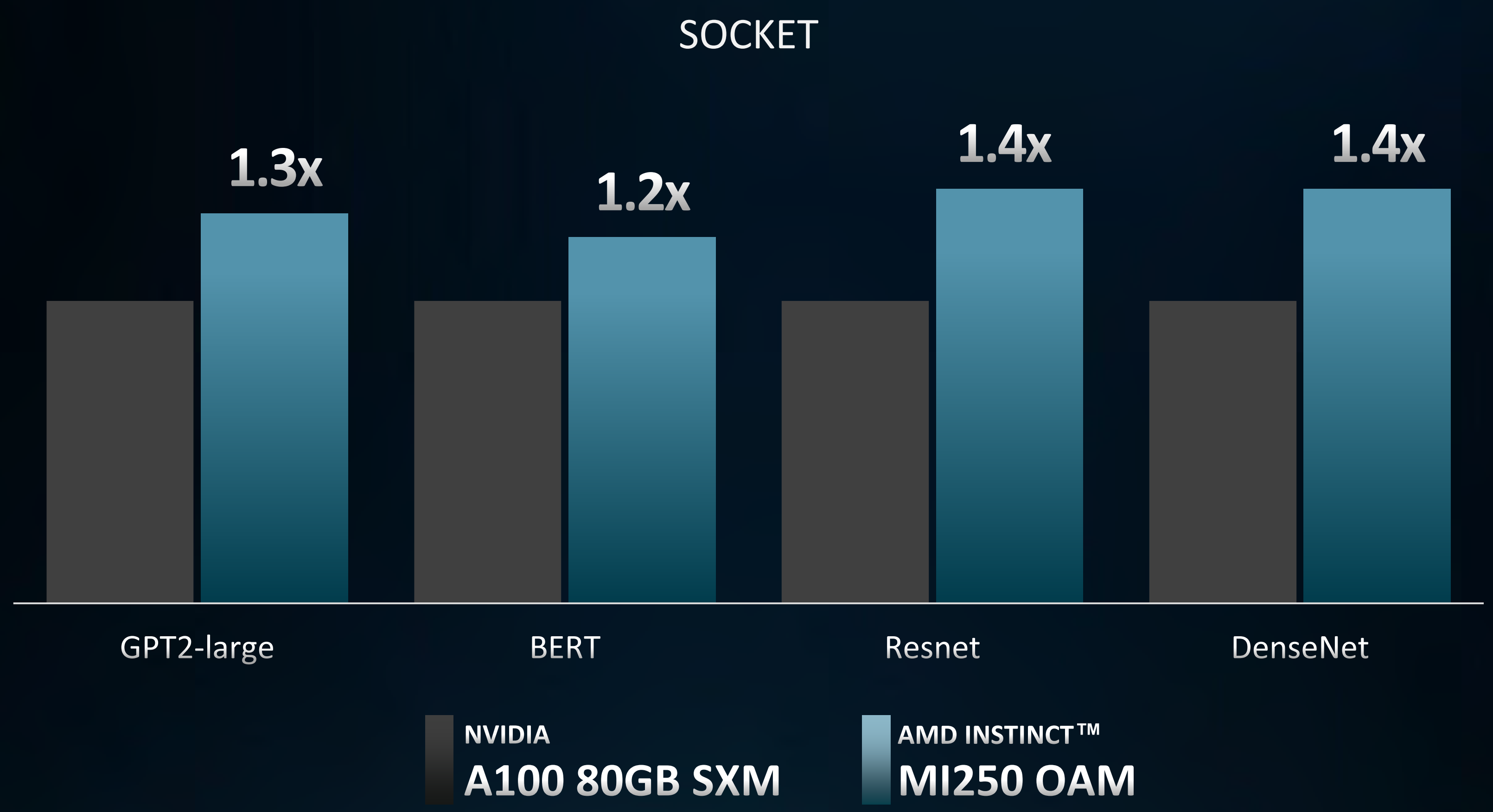
64GB HBM2e @ 1.6 TB/s



2-WAY & 4-WAY INFINITY
FABRIC™ BRIDGES

SR-IOV SUPPORT
FOR VMWARE®

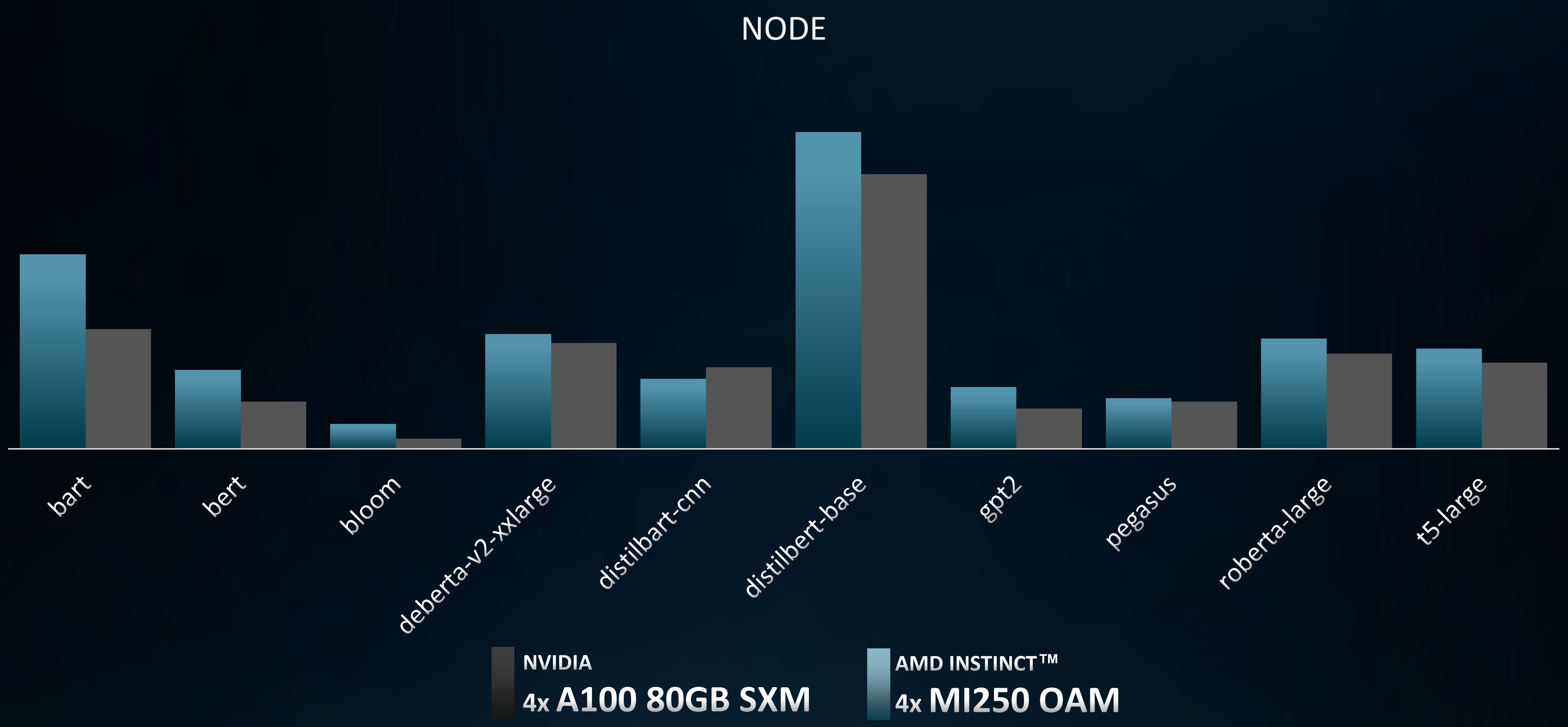
AMD INSTINCT™ MI250 GPU: LEADERSHIP PERFORMANCE



(8x) A100 SXM (80GB) w/ AMD EPYC 7742 VS (8x) MI250X OAM (128GB) w/optimized 2nd Gen EPYC
MI200-57, MI200-59, MI200-61, MI200-63

[Confidential - Distribution with NDA]

AMD INSTINCT™ MI250 GPU: LEADERSHIP PERFORMANCE



(4x) A100 SXM (80GB) w/ AMD EPYC 7742 VS (4x) MI250X OAM (128GB) w/optimized 2nd Gen EPYC
HUGGING FACE ITER/SEC

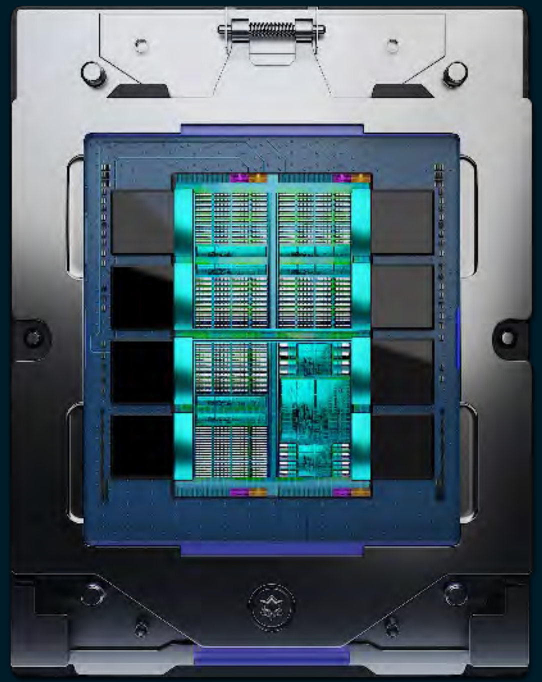
[Confidential - Distribution with NDA]

AMD INSTINCT™ MI300 OAM

THE WORLD'S FIRST DATA CENTER APU

- 4th Gen AMD Infinity Architecture: AMD CDNA™ 3 and EPYC™ CPU "Zen 4" Together
  CPU and GPU cores share a unified on-package pool of memory
- Groundbreaking 3D Packaging
  CPU | GPU | Cache | HBM
- Designed for Leadership Memory Capacity, Bandwidth and Application Latency
- APU Architecture Designed for Power Savings Compared to Discrete Implementation

Coming on 2H 2023



AMD CDNA 3 — Next-Gen Accelerator Architecture

ZEN 4 — 24 Leadership Data Center CPU cores

146B Transistors

128GB HBM3

See Endnote MI300-03. Preliminary data and projections, subject to change.

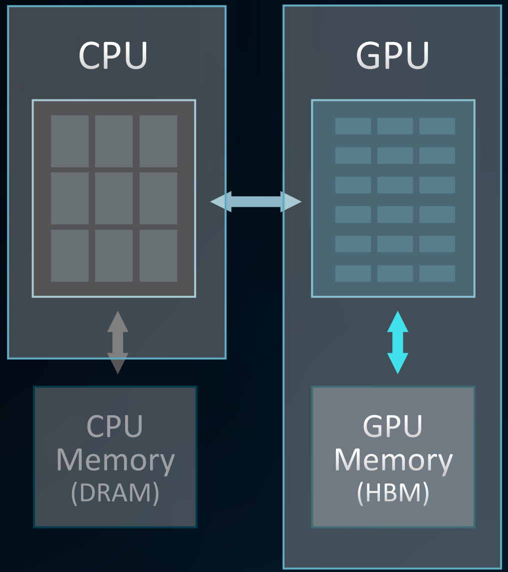
[Confidential - Distribution with NDA]

AMD INSTINCT™ MI300 OAM

UNIFIED MEMORY APU ARCHITECTURE BENEFITS

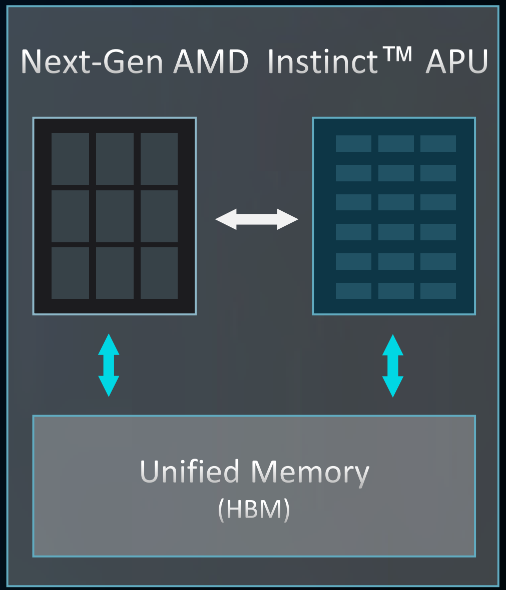
AMD CDNA™ 2 Coherent Memory Architecture

- Simplifies Programming
- Low Overhead 3rd Gen Infinity Interconnect
- Industry Standard Modular Design



AMD CDNA™ 3 Unified Memory APU Architecture

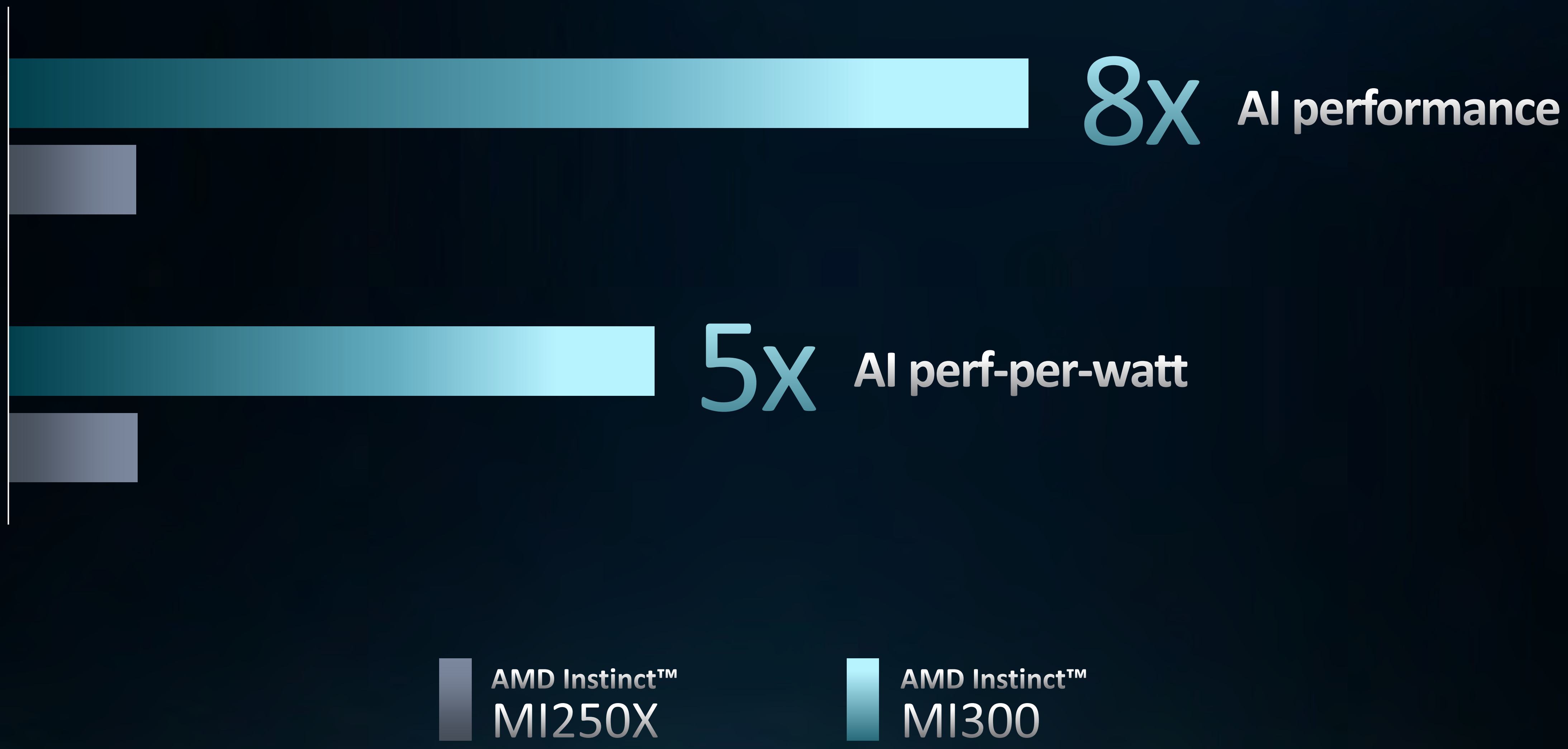
- Eliminates Redundant Memory Copies
- High-Efficiency 4th Gen AMD Infinity Architecture
- Low TCO with Unified Memory APU Package



*AMD Projections

[Confidential - Distribution with NDA]

AMD INSTINCT™ MI300 OAM



SOFTWARE

Building the ROCm ecosystem

together we advance_

RADEON OPEN COMPUTE

AMD ROCm

- Unlocked GPU Performance to Accelerate Computational Tasks
- Optimized for HPC and AI Workloads at Scale
- Open Source Enabling Collaboration, Innovation, and Differentiation

AI Framework Support	ONNX Runtime	PyTorch	TensorFlow	Jax		
Programming Models	OpenMP API	HIP API	OpenCL™			
Libraries	BLAS	RAND	FFT	MIGraphX	MIVisionX	PRIM
	SOLVER	ALUTION	SPARSE	THRUST	MIOpen	RCCL
Compilers & Tools	Compiler	Profiler	Tracer	Debugger	hipify	TENSILE
Drivers & Runtime	RedHat, CentOS, SLES & Ubuntu Device Drivers and Run-Time					
Deployment Tools	ROCm Validation Suite	ROCm Data Center Tool	ROCm SMI			

[Confidential - Distribution with NDA]

"DROP-IN" SUPPORT FOR MAJOR ML FRAMEWORKS

	Source	Support
TensorFlow	TensorFlow GitHub	Fully upstream with weekly synchronization & CI
PyTorch	PyTorch GitHub	Fully upstream with stable release on pytorch.org, ~50 servers running daily CI
ONNX RUNTIME	ONNX-RT GitHub	MIGraphX/HIP EPs fully upstream in MSFT CI
JAX	GitHub public fork	Initial enablement in AMD hosted public repo
DeepSpeed	DeepSpeed GitHub	Fully upstream in MSFT CI
Triton	Triton GitHub	Development support upstream
CuPy	cupy.dev	PIP binary package with ~70% functional coverage

INNOVATING THROUGH TECHNOLOGY PARTNERSHIPS

Microsoft Azure

Accelerating ROCm Performance on AI

"We're also deepening our investment in the open-source PyTorch framework, working with the PyTorch Core team and AMD both to **optimize the performance and developer experience for customers running PyTorch on Azure, and to ensure that developers' PyTorch projects work great on AMD hardware.**"

Kevin Scott
Executive Vice President and CTO, Microsoft

PyTorch

Optimizing ROCm for PyTorch

"We are excited to partner with AMD to grow our PyTorch support on ROCm, enabling the vibrant PyTorch community to adopt the latest generation of AMD Instinct GPUs quicker than ever with great performance for major AI use cases running on PyTorch."

Soumith Chintala
Co-Creator and Lead of PyTorch, Meta AI

LANDING AI

Developing Data Centric AI for Instinct™ GPUs

"We are excited to partner with AMD to leverage AMD MI200 GPUs and the ROCm stack to port and optimize LandingLens, our GPU-optimized computer vision software application. The power of AMD GPUs and the maturity of ROCm will help Landing AI continue to deliver acceleration of high-resolution computer vision models with the purpose of providing better insights into manufacturing defects."

Andrew Ng
CEO of Landing AI and Adjunct Professor, Stanford University

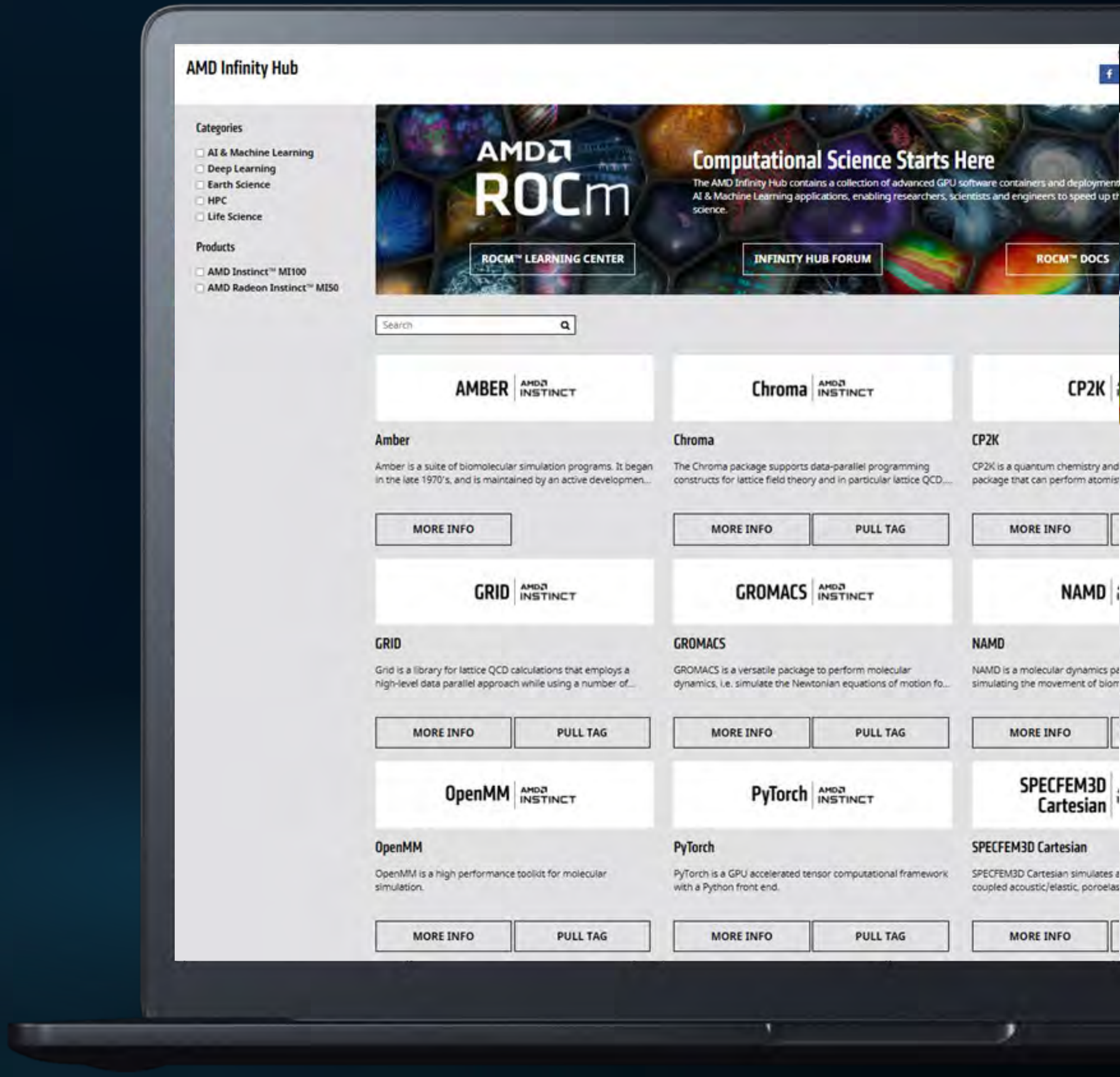
[Confidential - Distribution with NDA]

AMD INFINITY HUB CONTAINERS

MI200 SUPPORT
Both PCIe and OAM versions

MOST POPULAR FRAMEWORKS AVAILABLE
TensorFlow and PyTorch

FURTHER ML OPTIONS PLANNED
BERT, ResNet50 etc. etc.

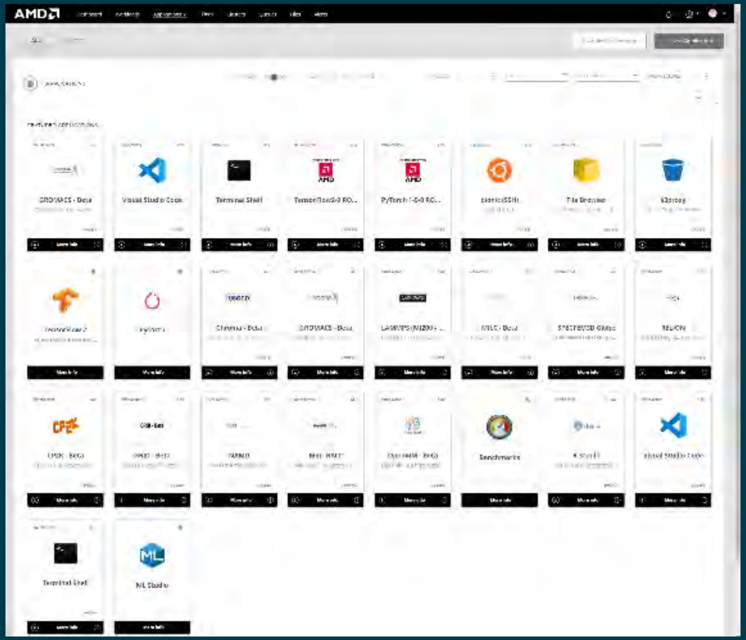


AMD.com/InfinityHub

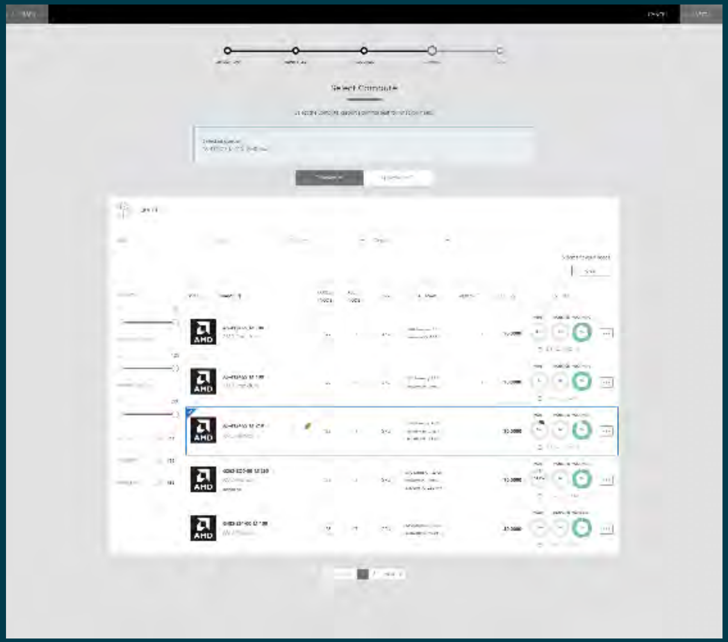
[Confidential - Distribution with NDA]

AMD ACCELERATOR CLOUD

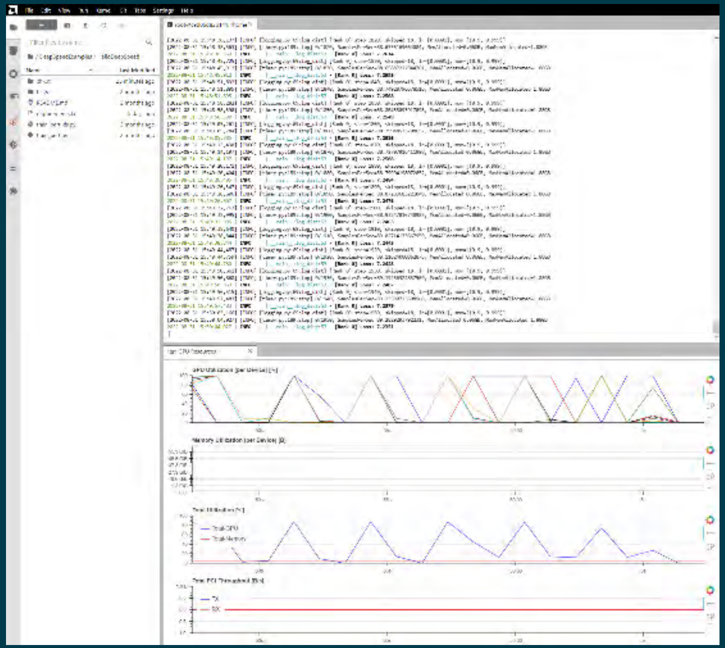
STEP 1



STEP 2



STEP 3

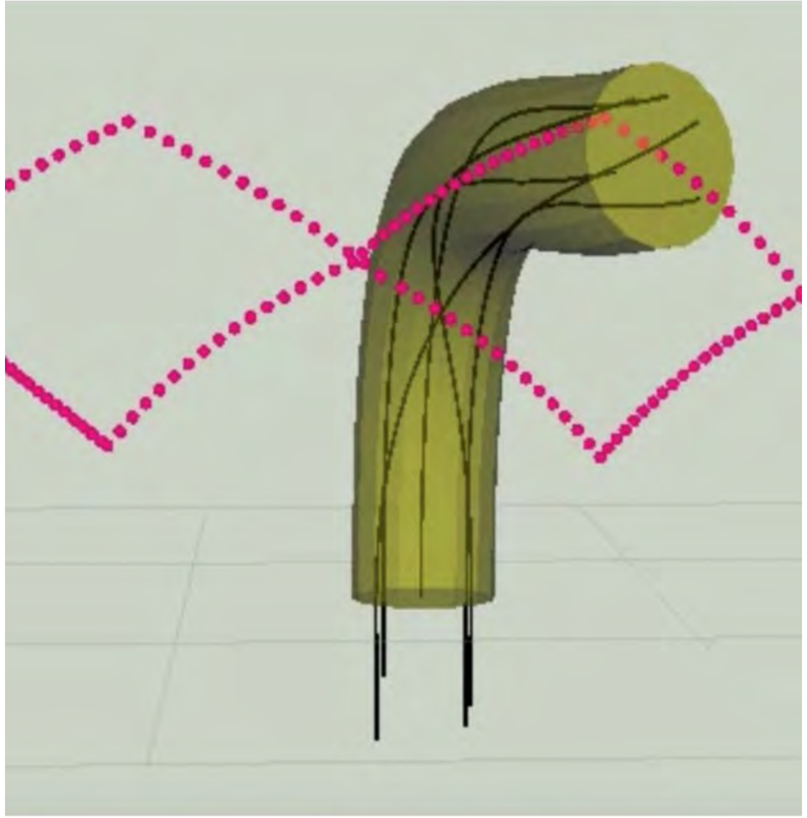
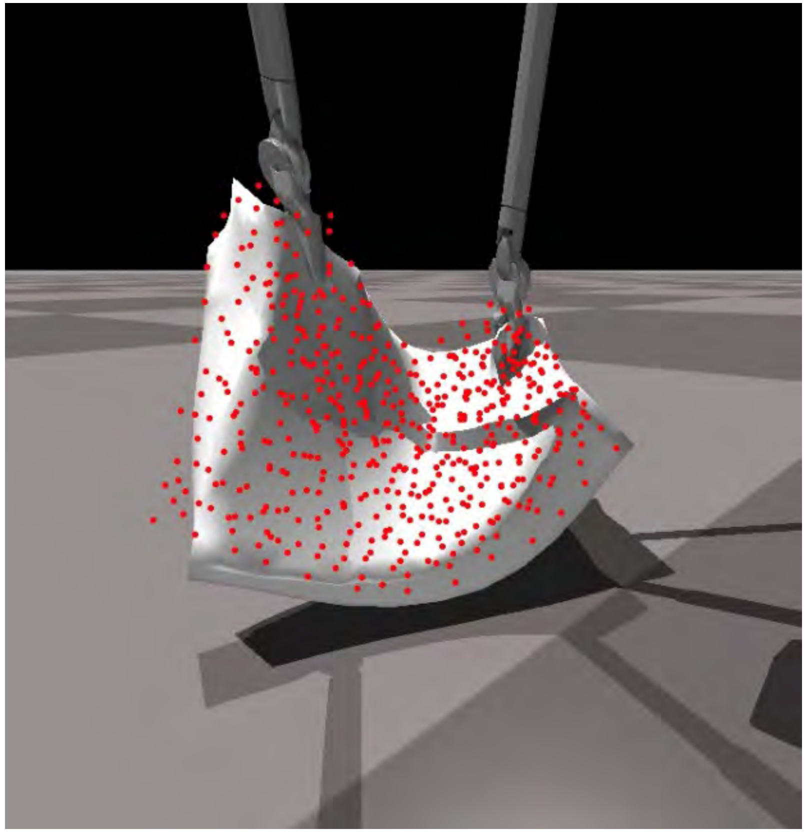
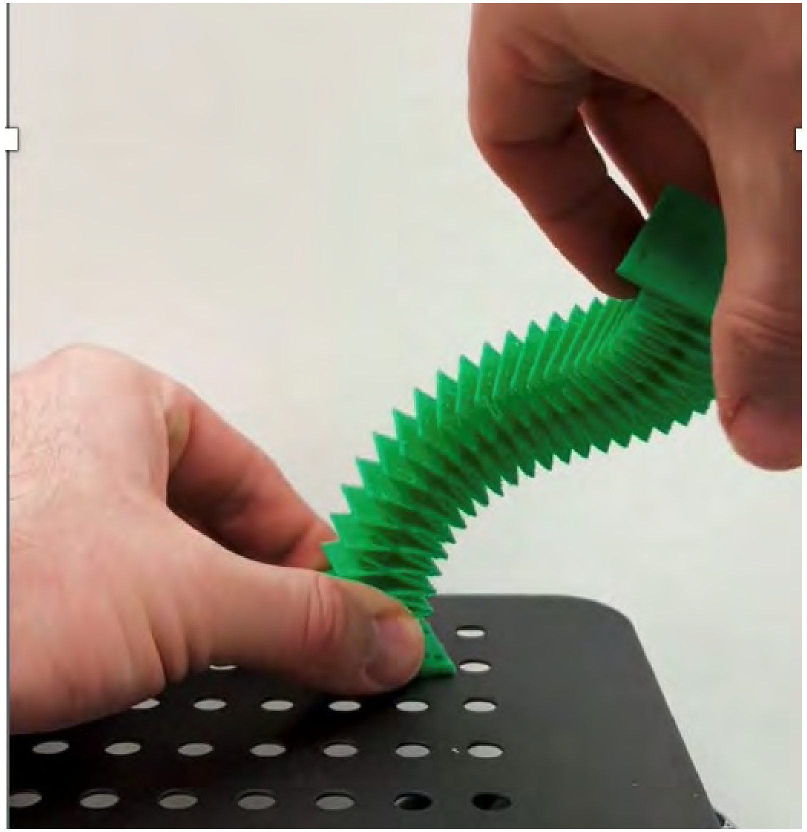


US-BASED FREE-TRIAL ENVIRONMENT

acc.amd.com

THANK YOU

ian.ferreira@amd.com

Tentacle-Like Continuum Robots and Sub-Task Automation

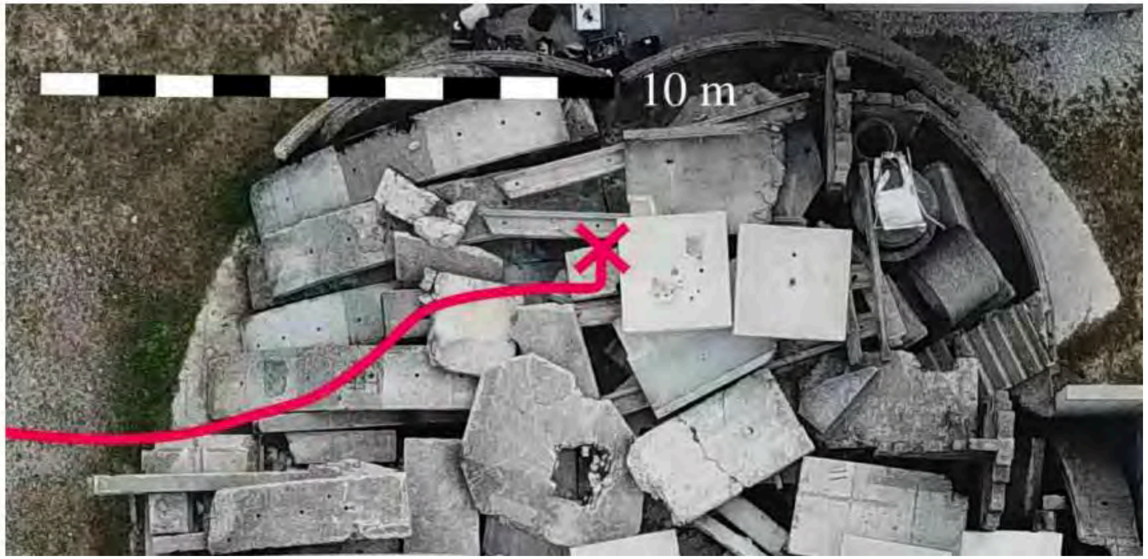
Alan Kuntz (alan.kuntz@utah.edu)
Assistant Professor, Kahlert School of Computing and Robotics Center
University of Utah

Continuum Manipulators

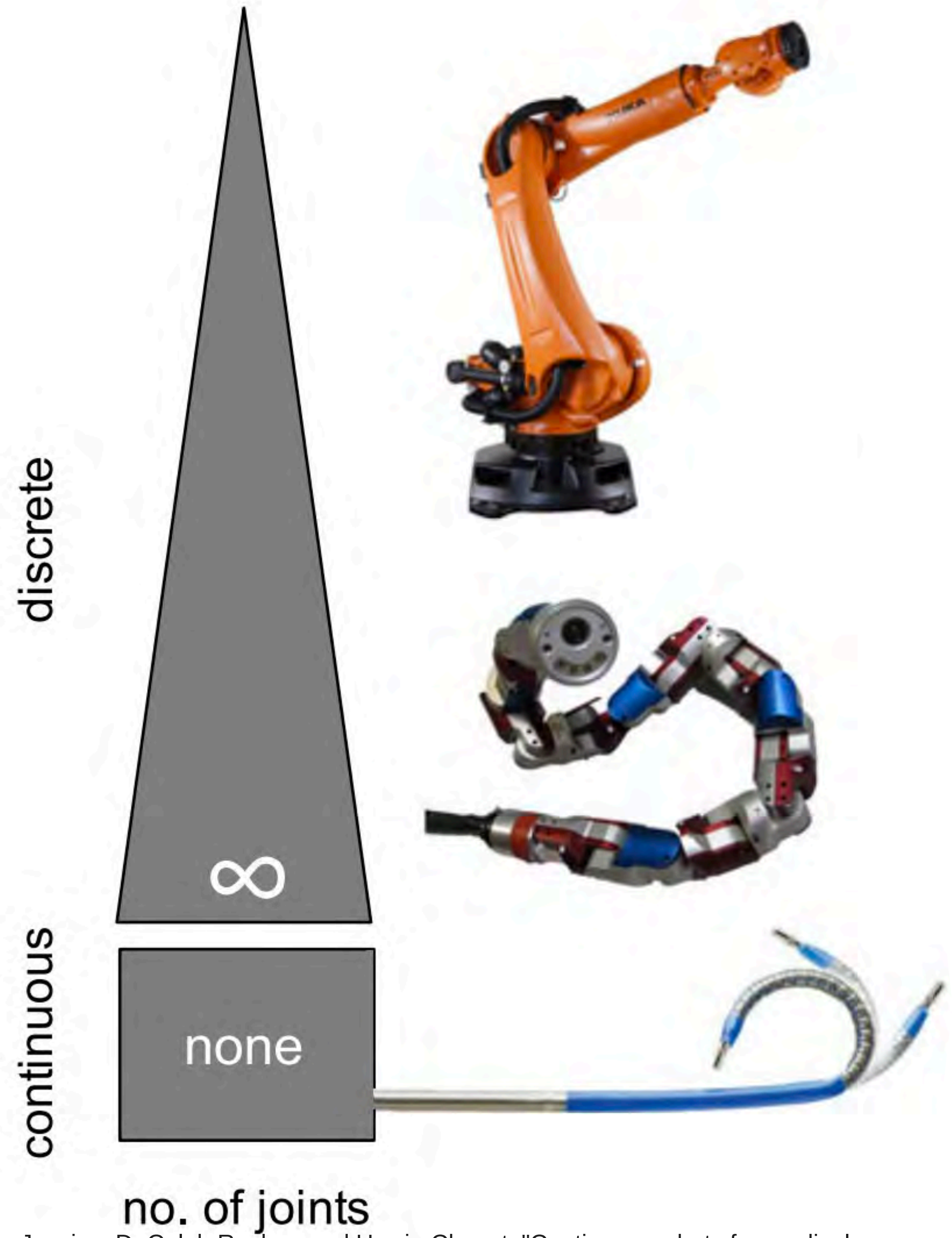
Curved, compliant manipulators

Snake through constrained, complex spaces

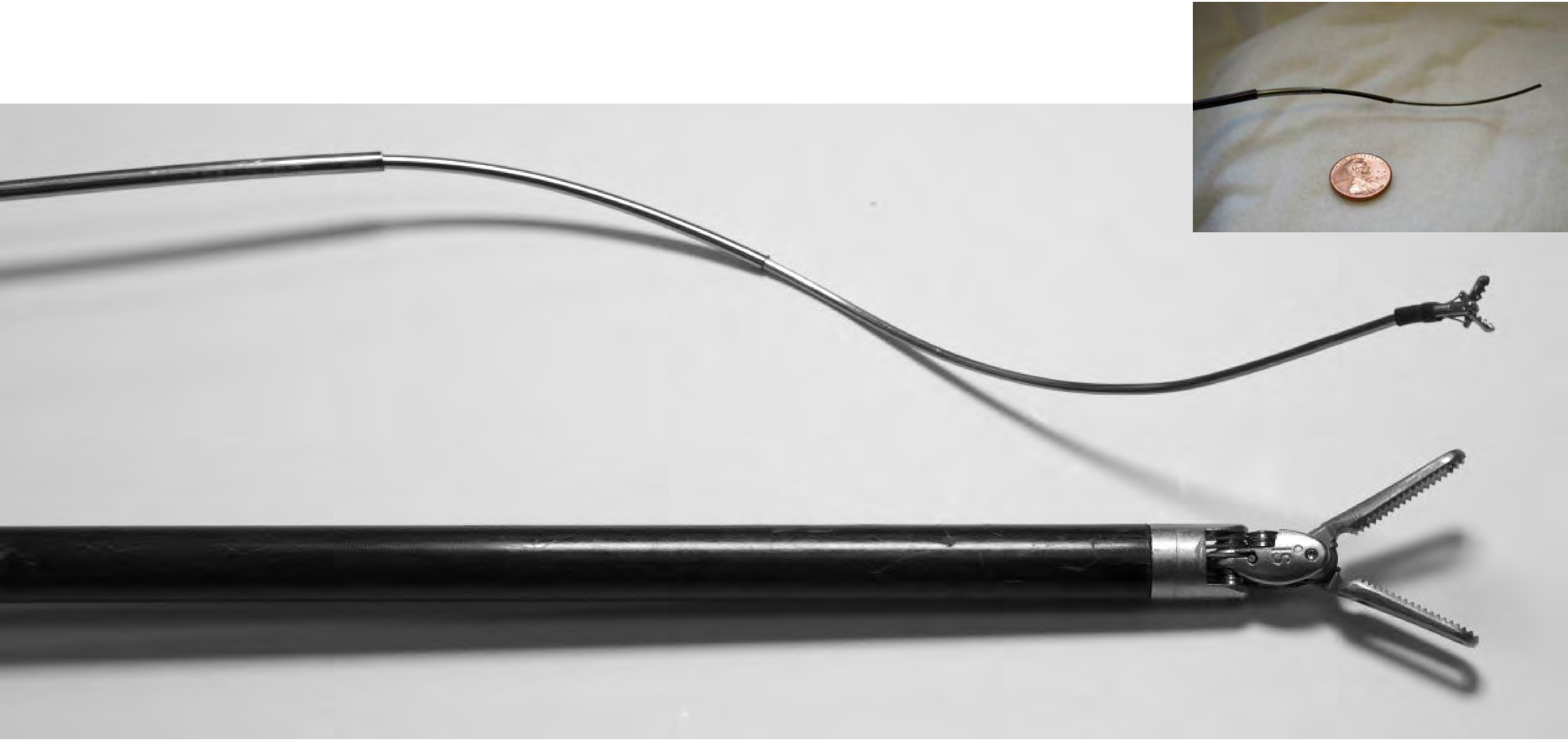
Soft interaction with the environment



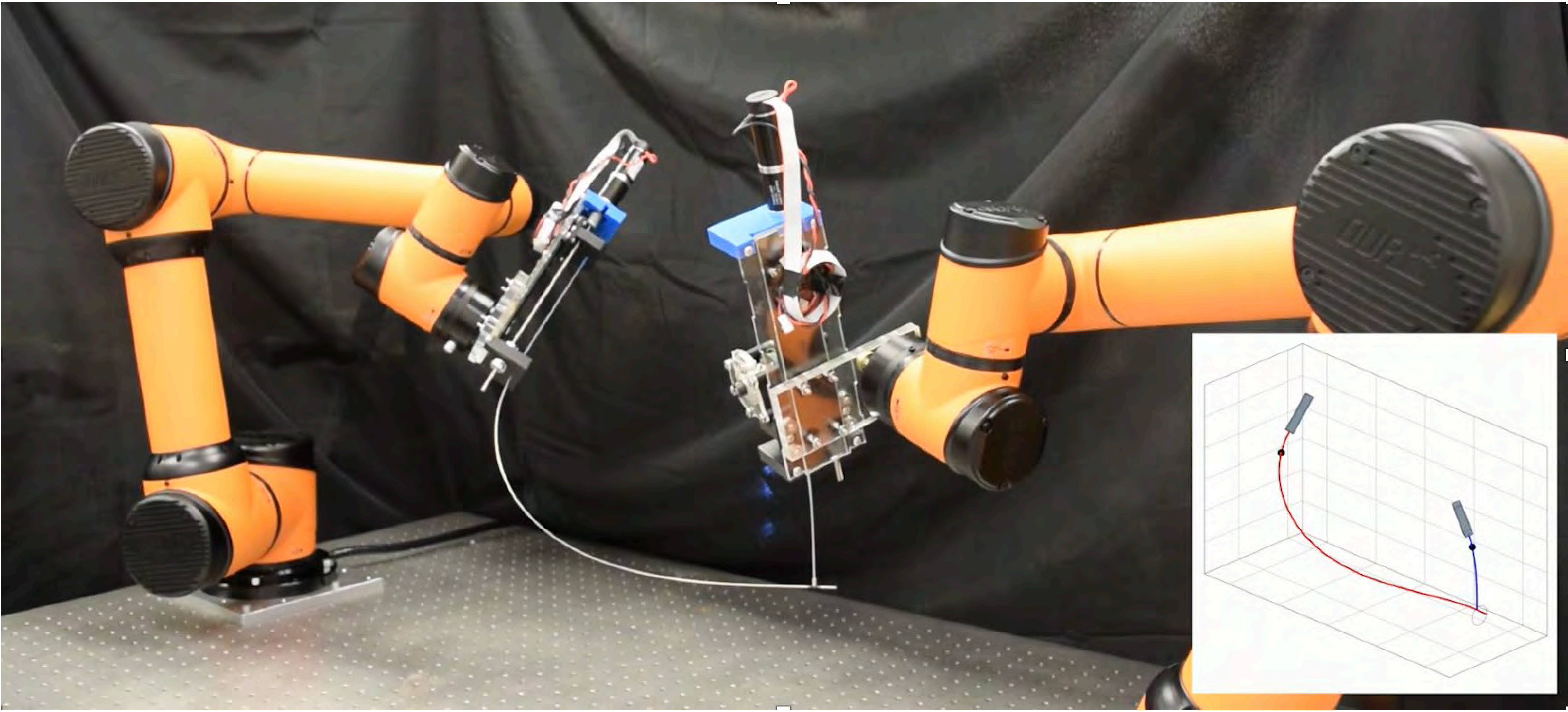
der Maur, Pascal Auf, et al. "RoBoa: Construction and evaluation of a steerable vine robot for search and rescue applications." *2021 IEEE 4th International Conference on Soft Robotics (RoboSoft)*. IEEE, 2021

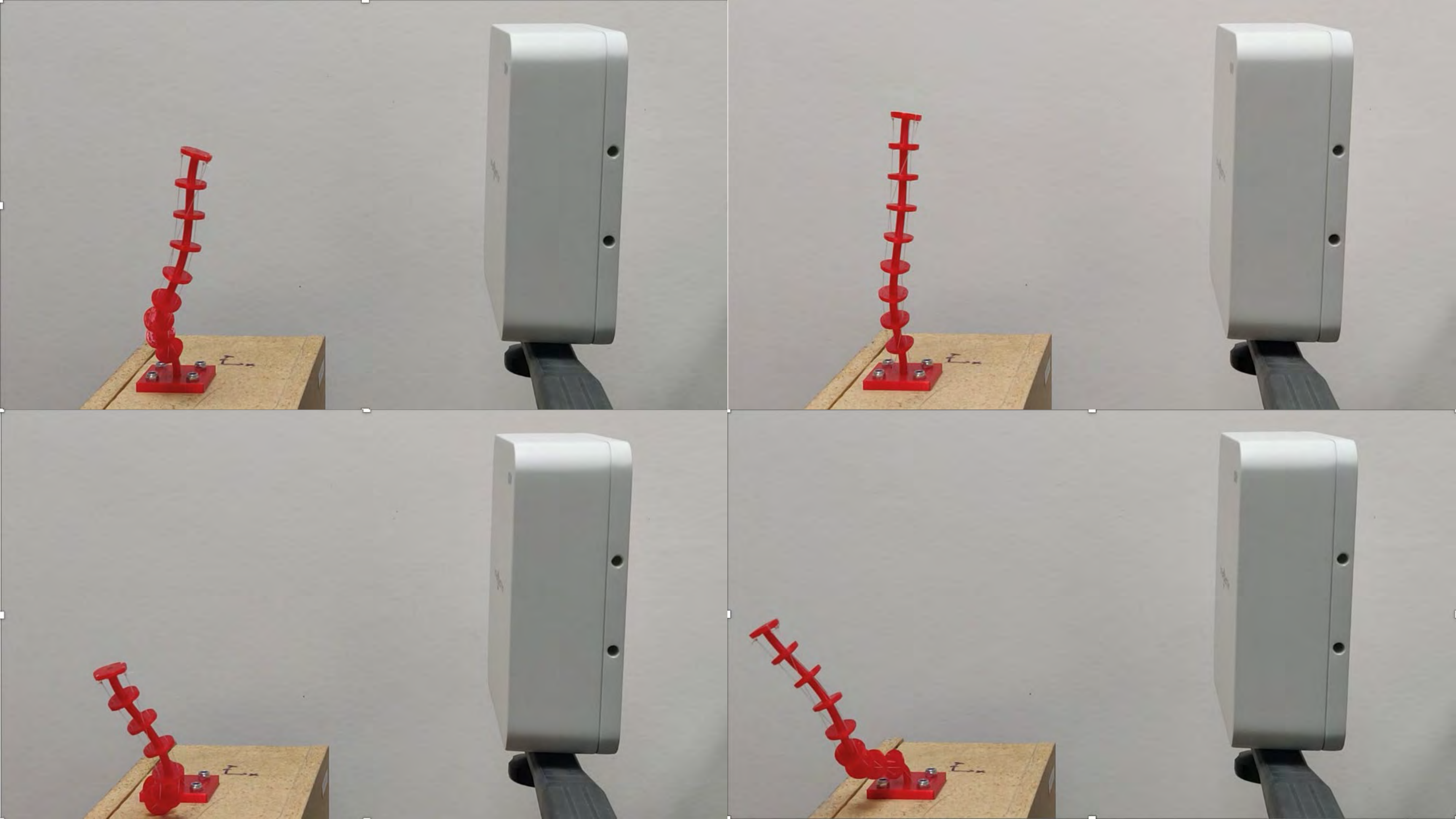


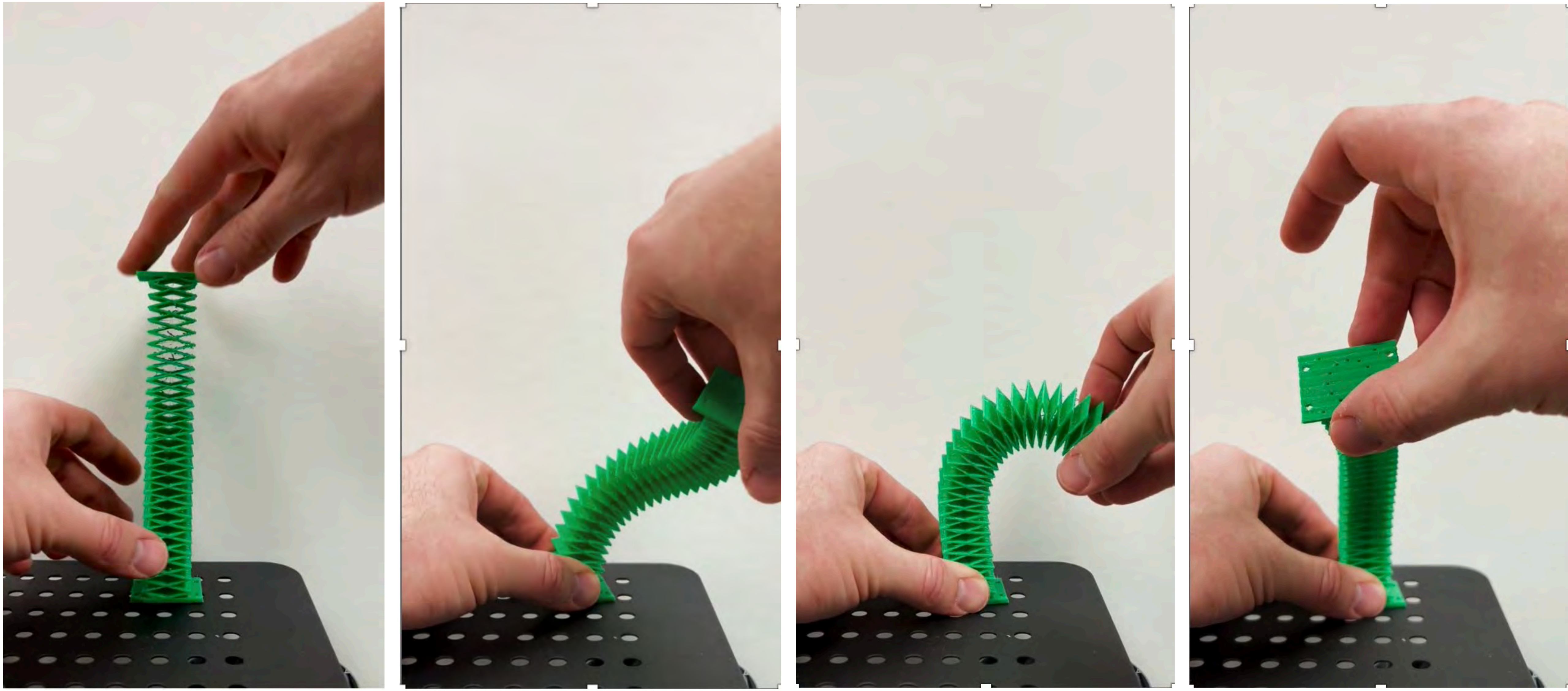
Burgner-Kahrs, Jessica, D. Caleb Rucker, and Howie Choset. "Continuum robots for medical applications: A survey." *IEEE Transactions on Robotics* 31.6 (2015): 1261-1280.

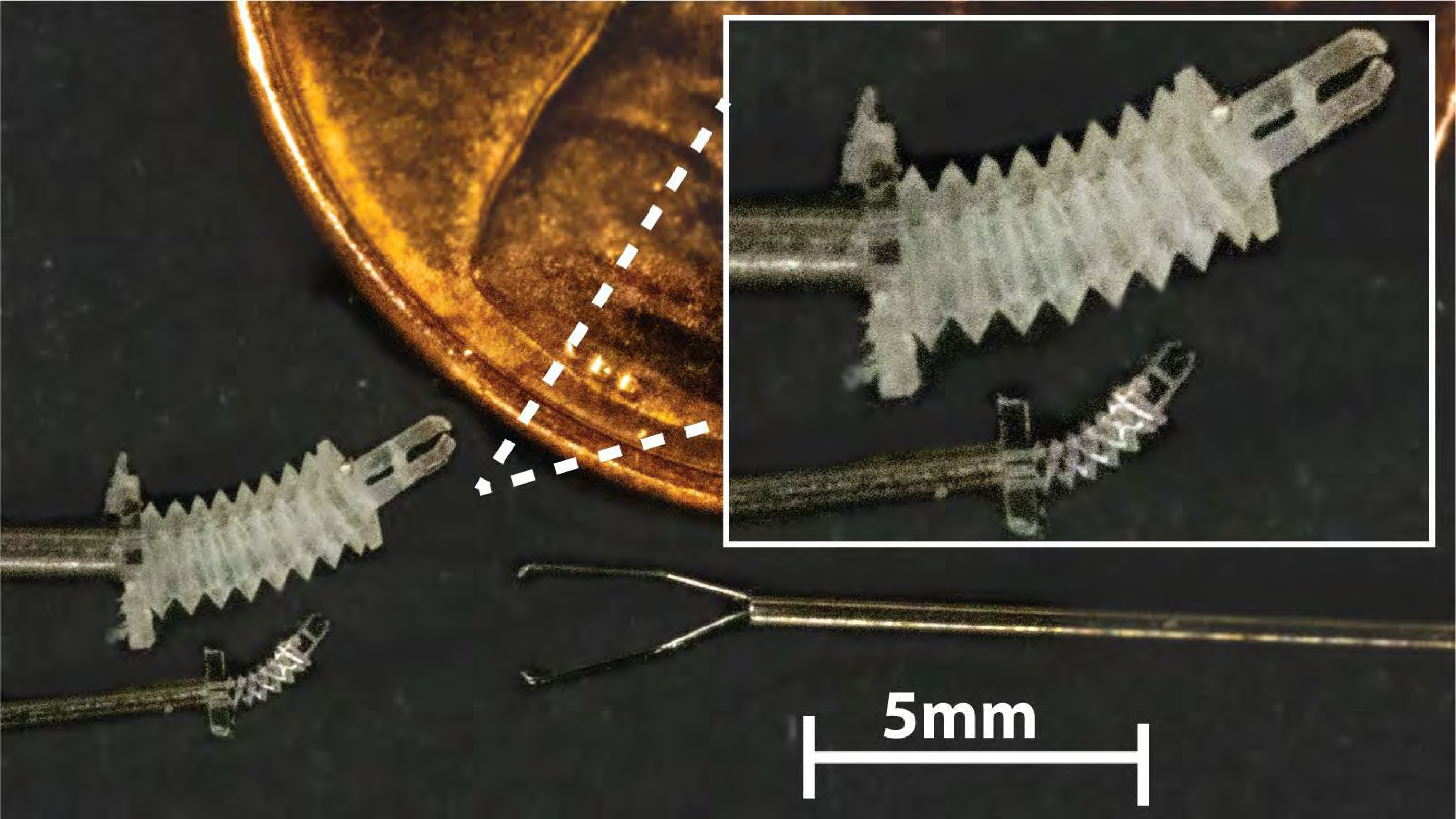


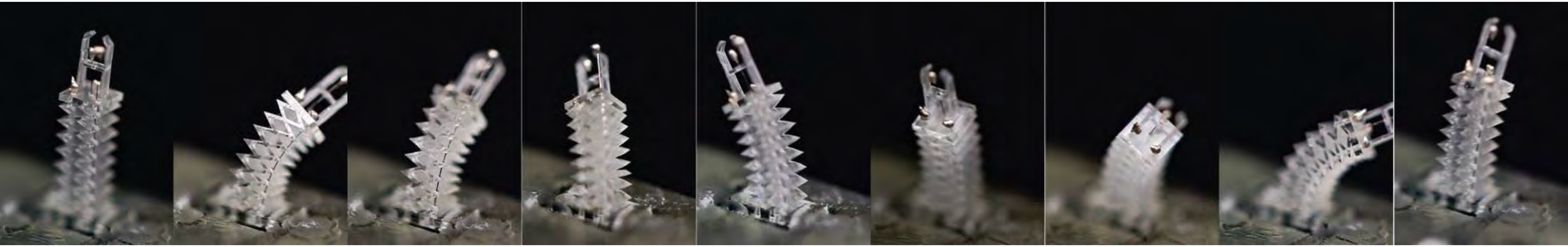
Gilbert, et.al. "Concentric tube robots: The state of the art and future directions." *Robotics Research* (2016): 253-269.



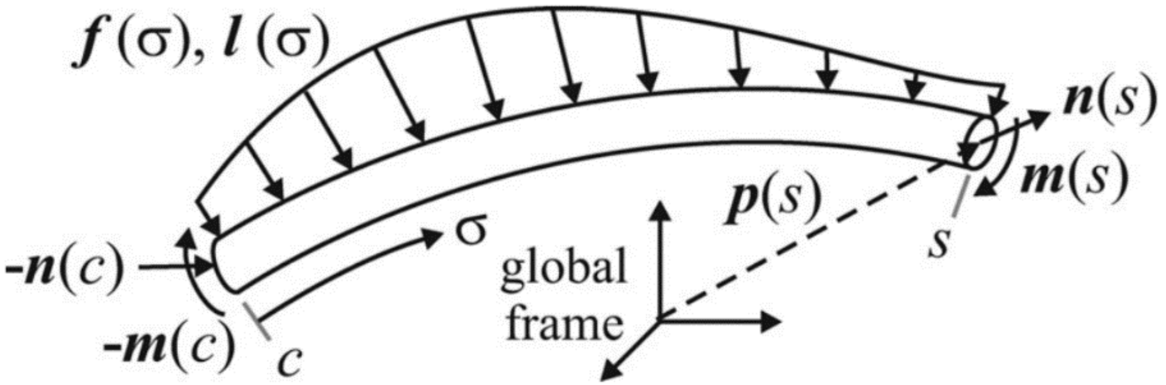





5mm

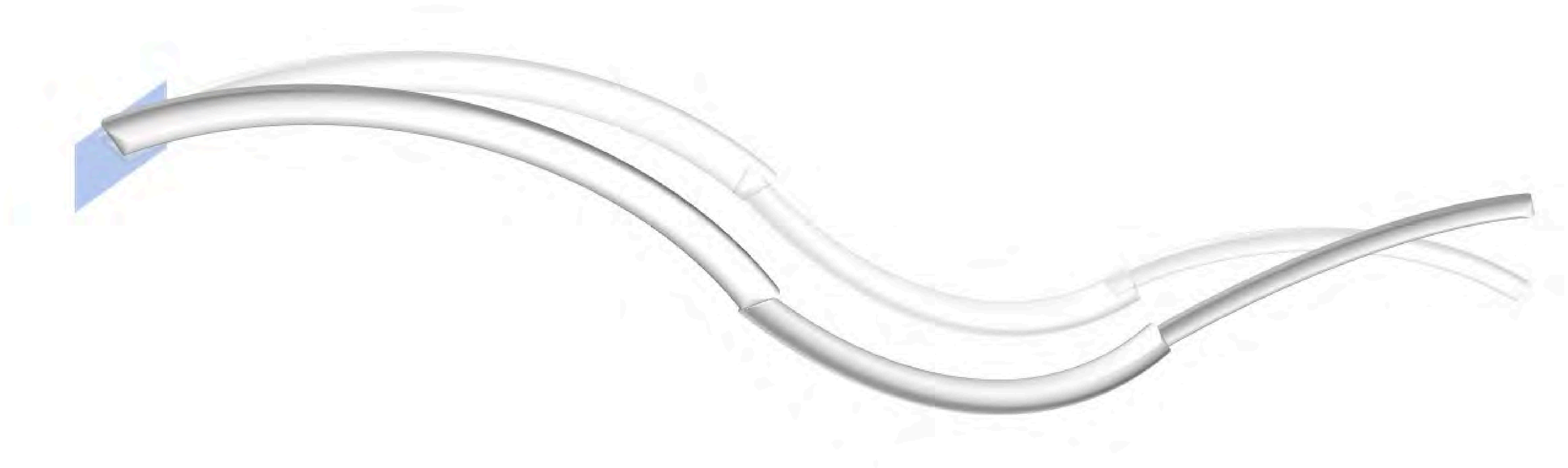


Physics-based shape model

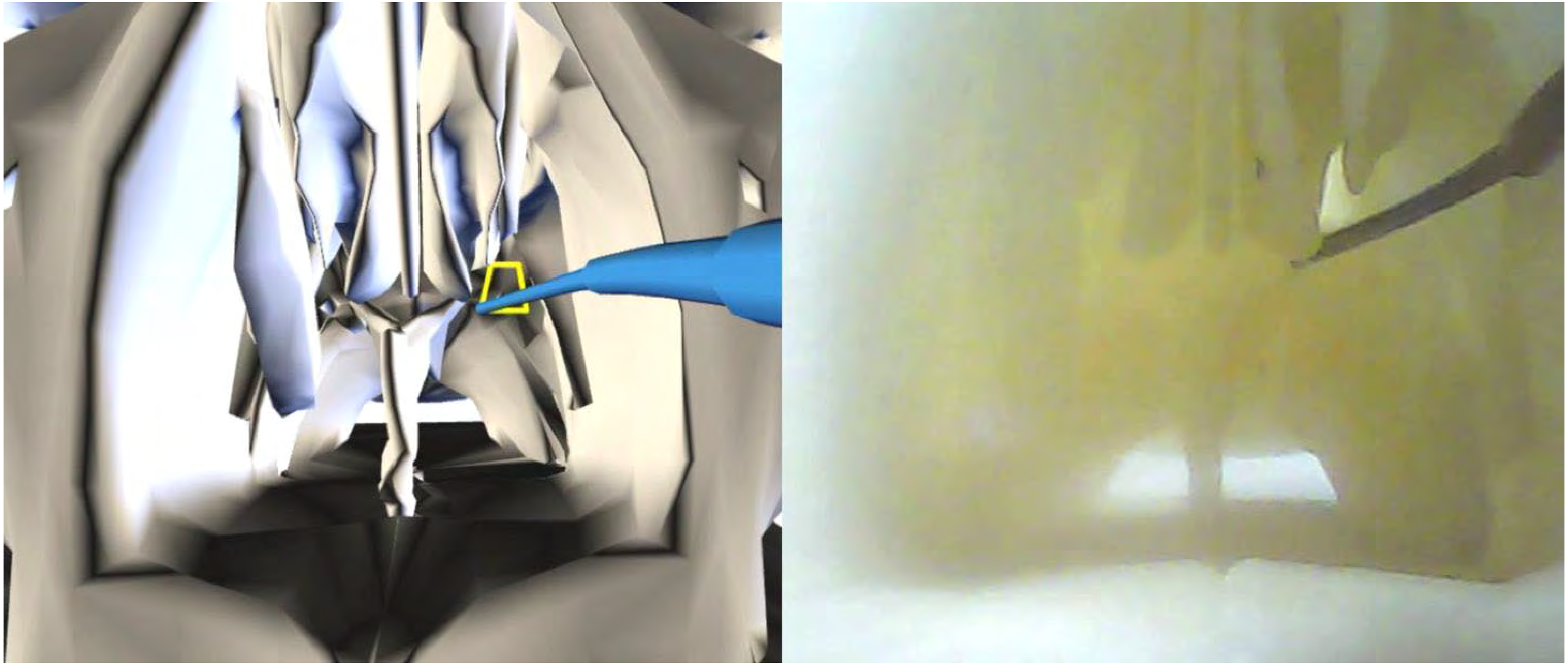


Rucker et al. 2011 (TRO)

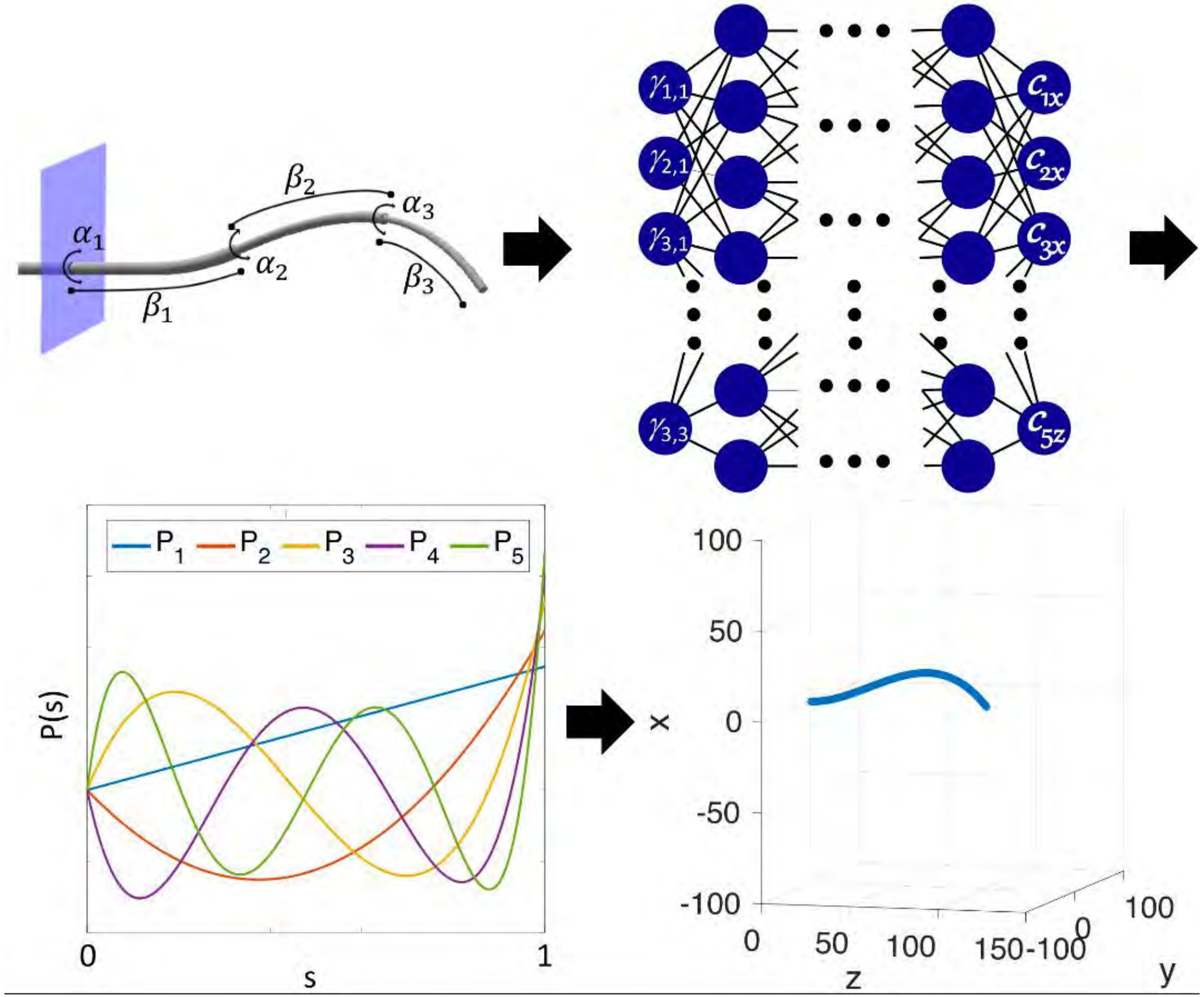
Shape model is frequently inaccurate



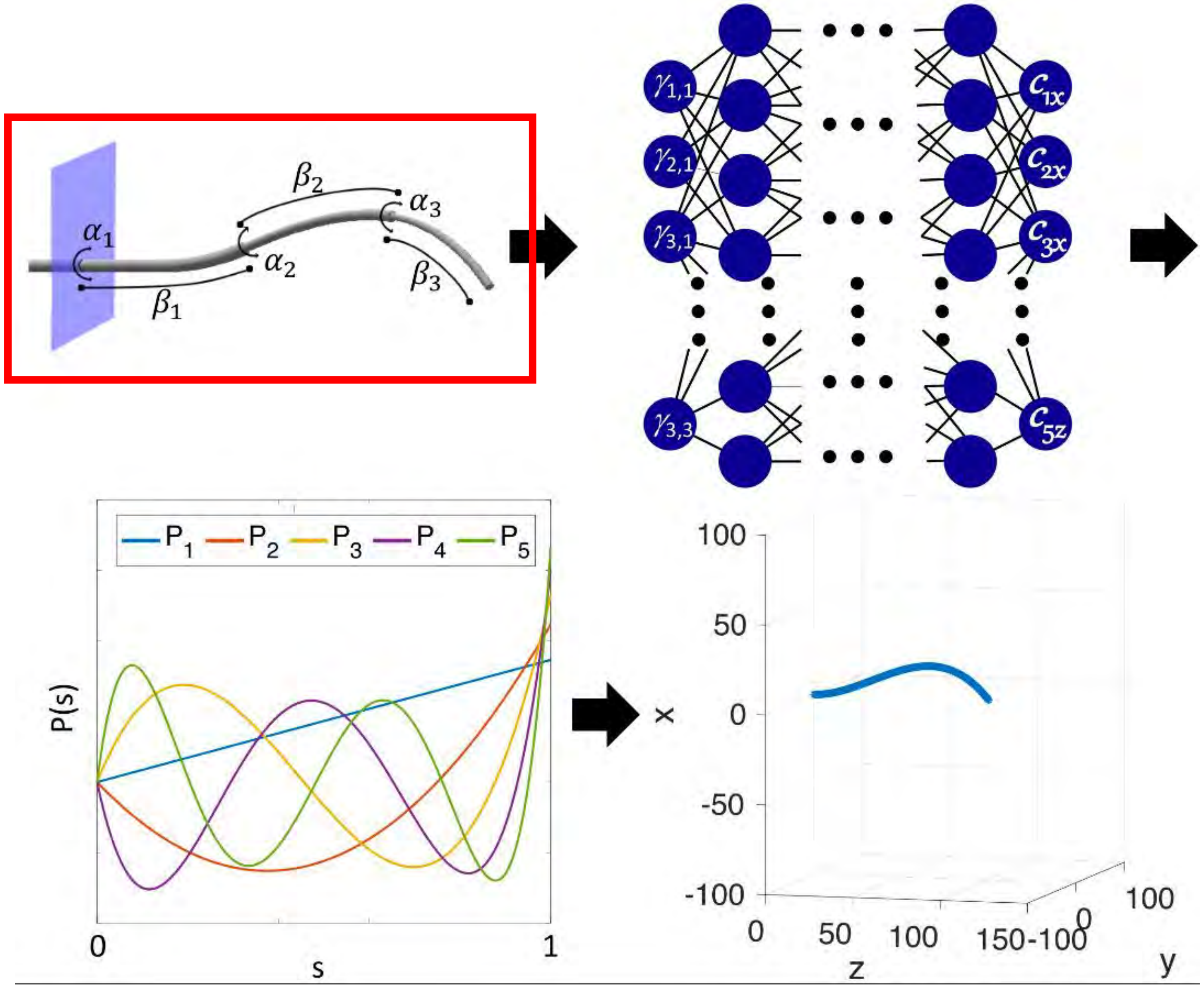
Unintended collisions



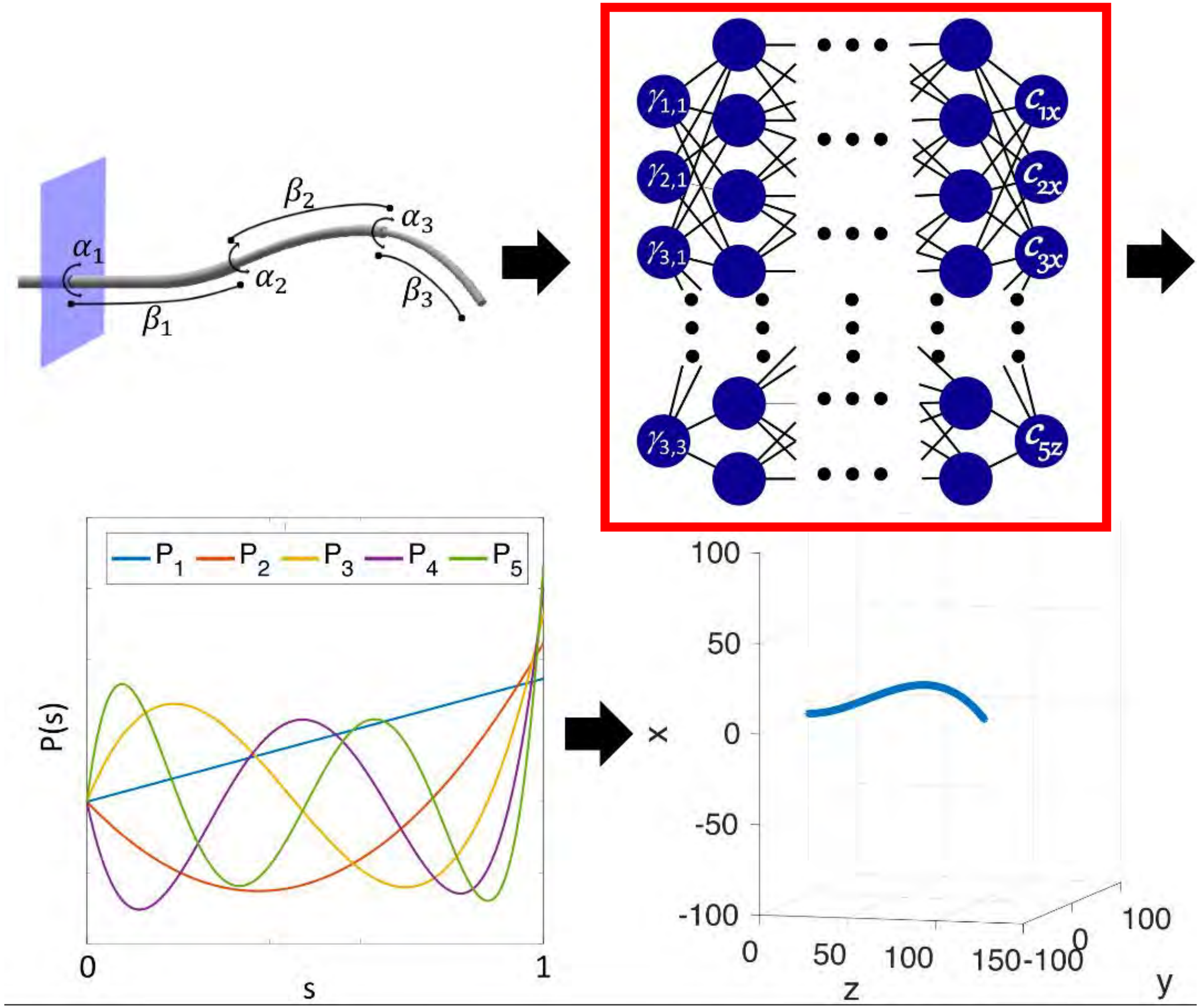
Learning a shape model



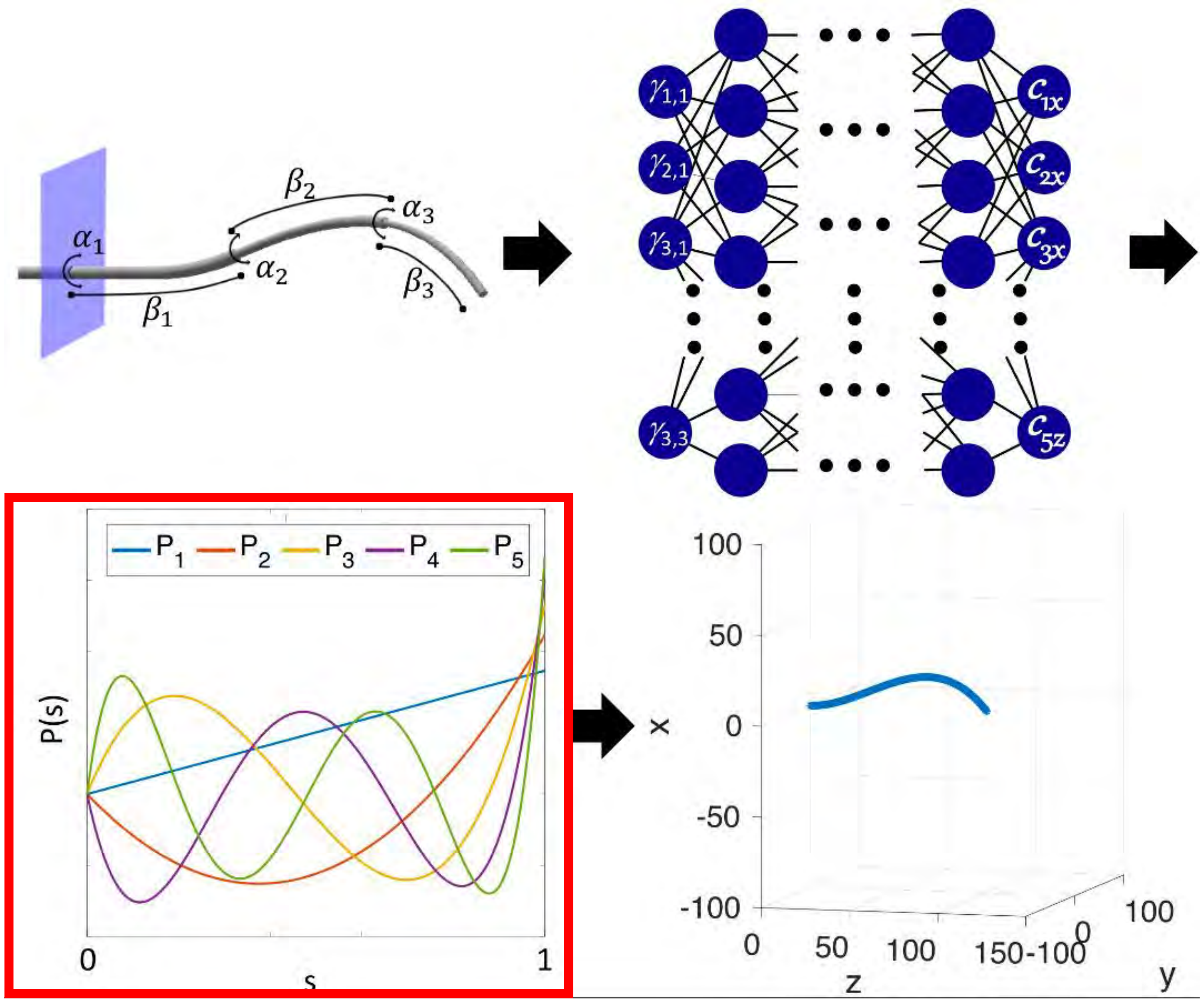
Learning a shape model



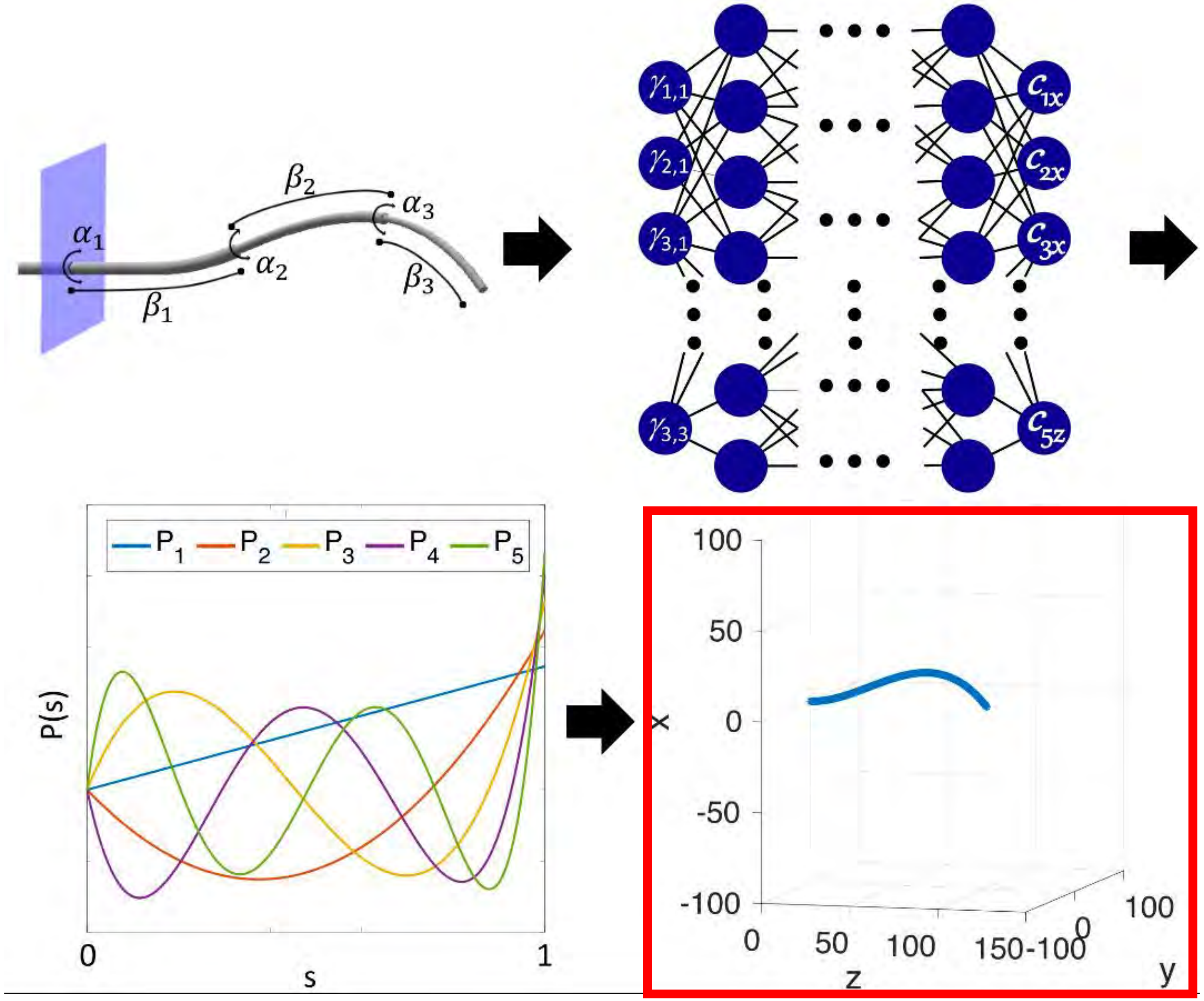
Learning a shape model



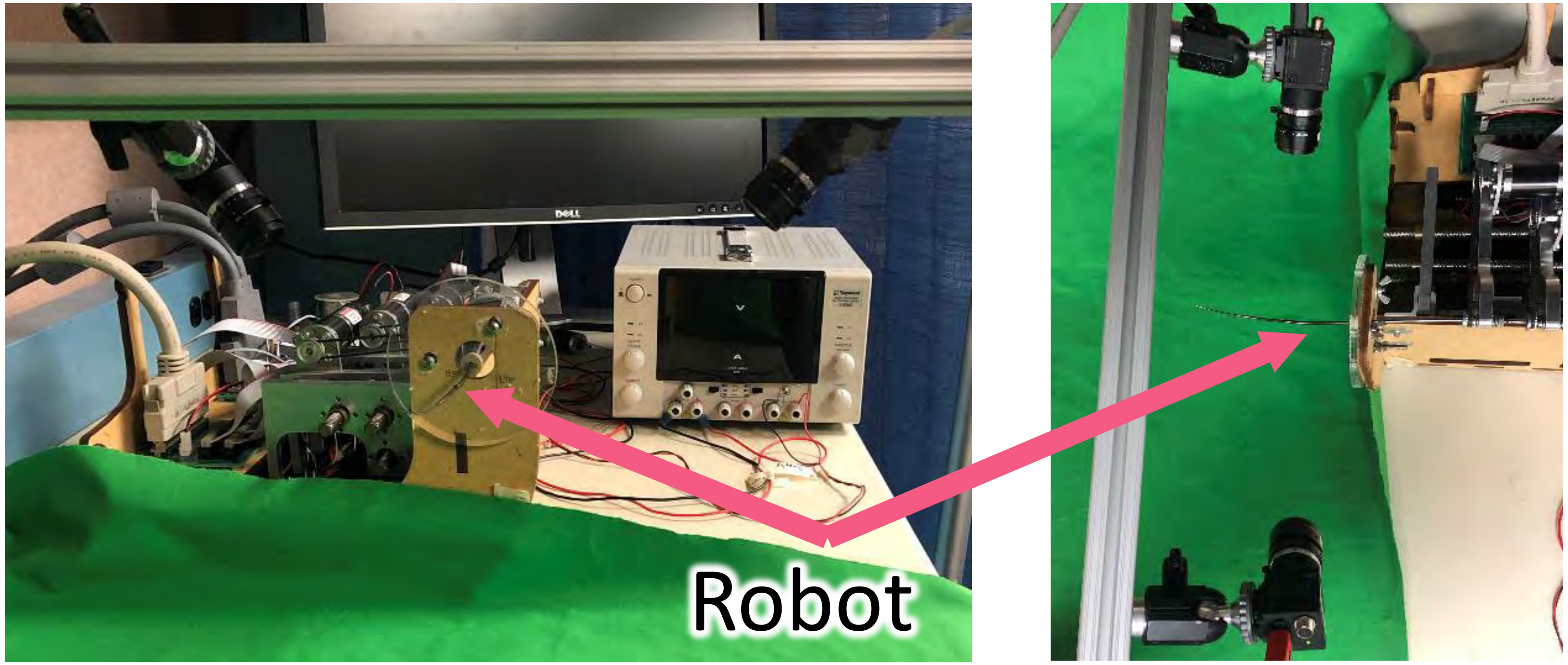
Learning a shape model



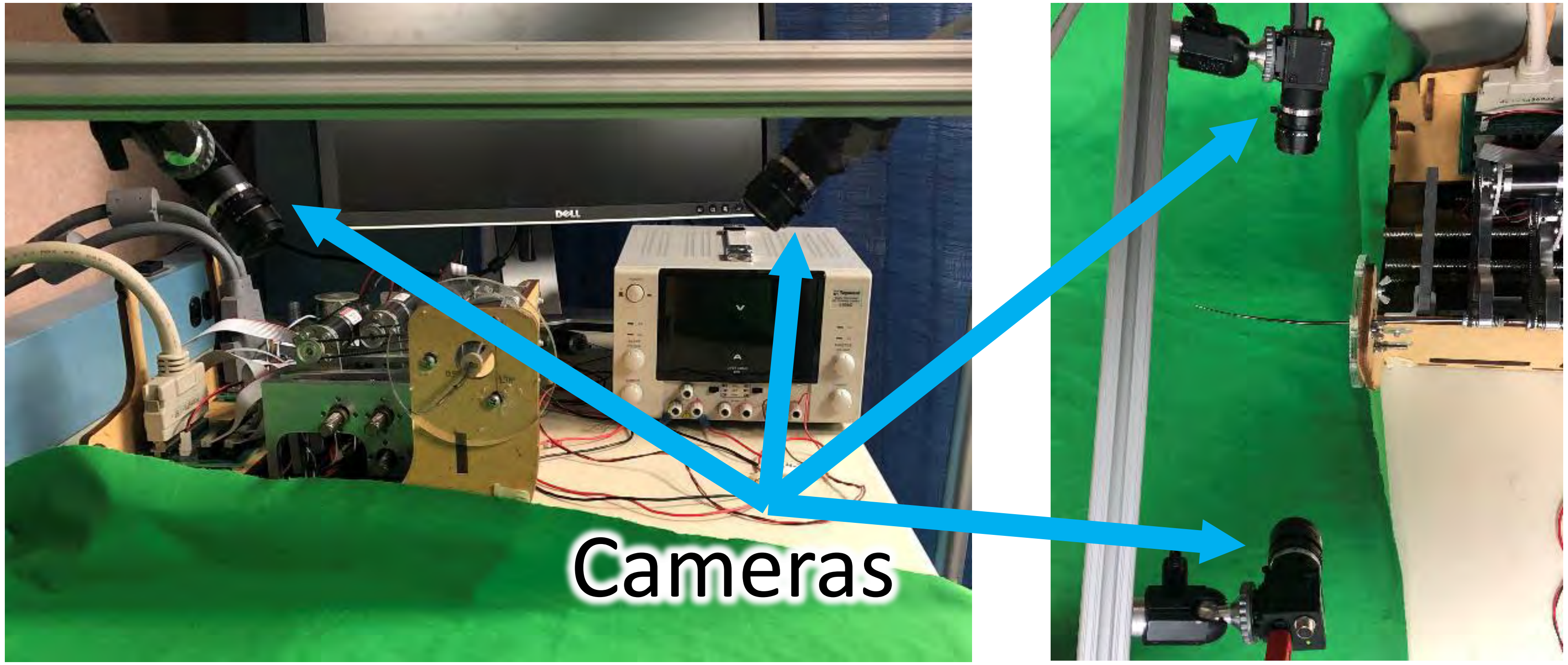
Learning a shape model



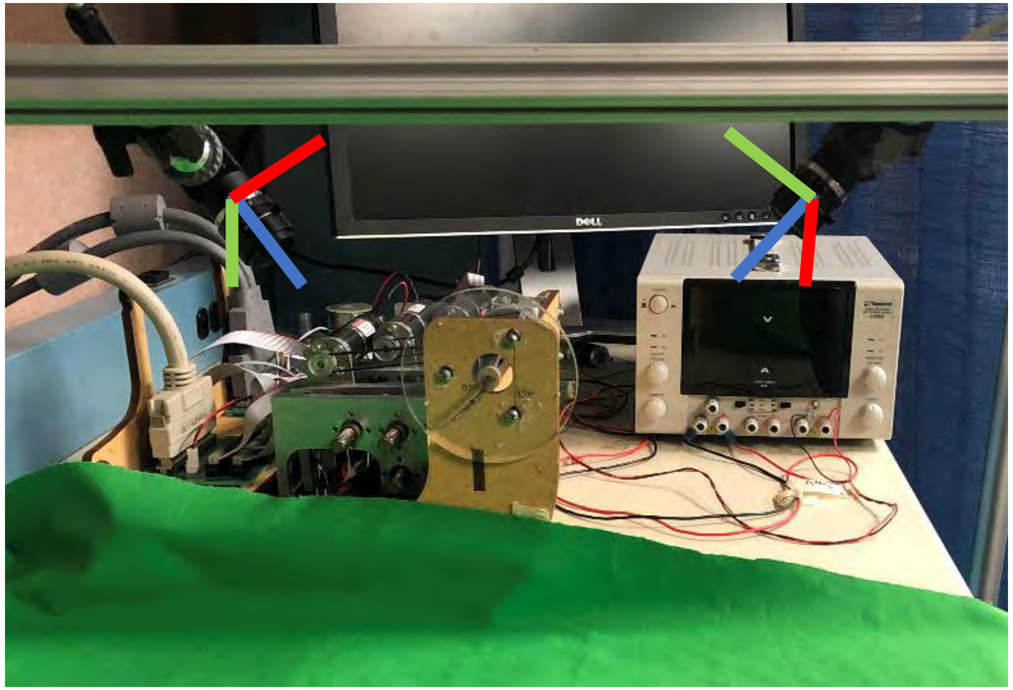
Training data: ground truth from robot

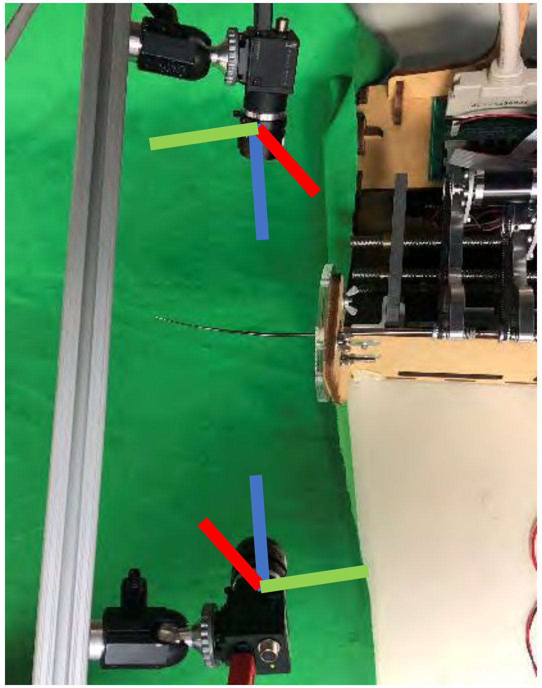


Training data: cameras with known location

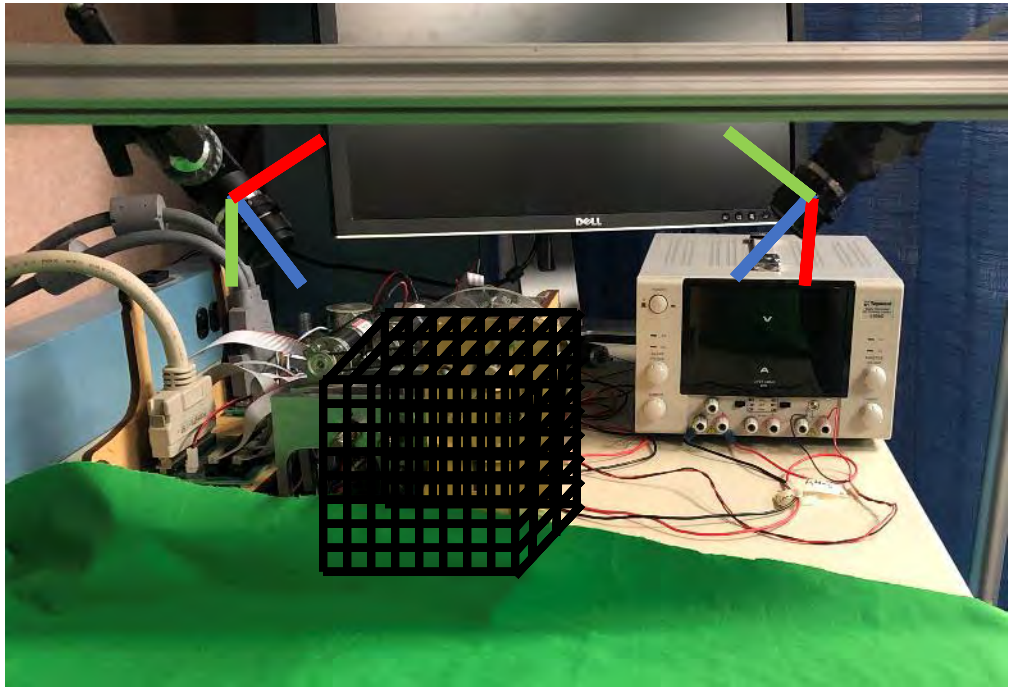


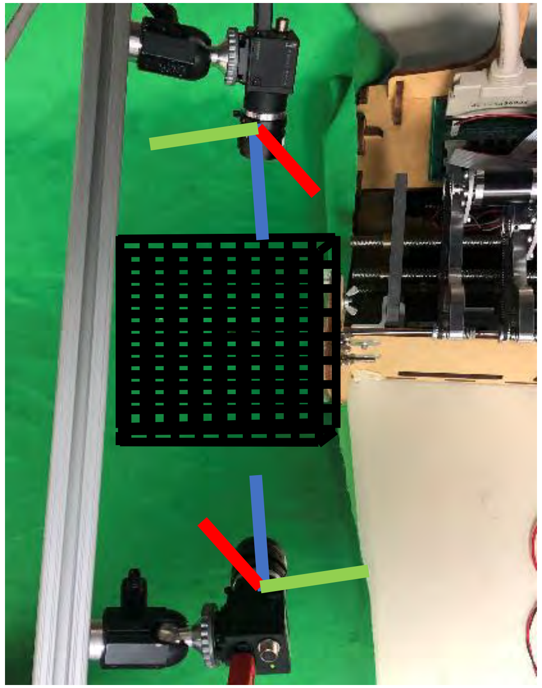
Training data: cameras with known location



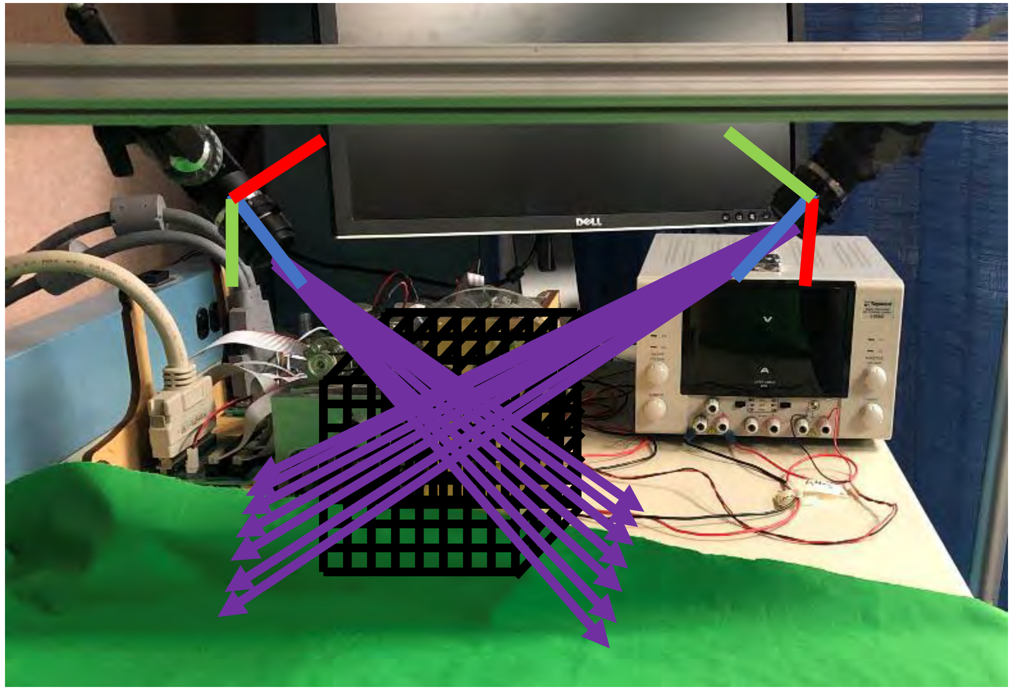


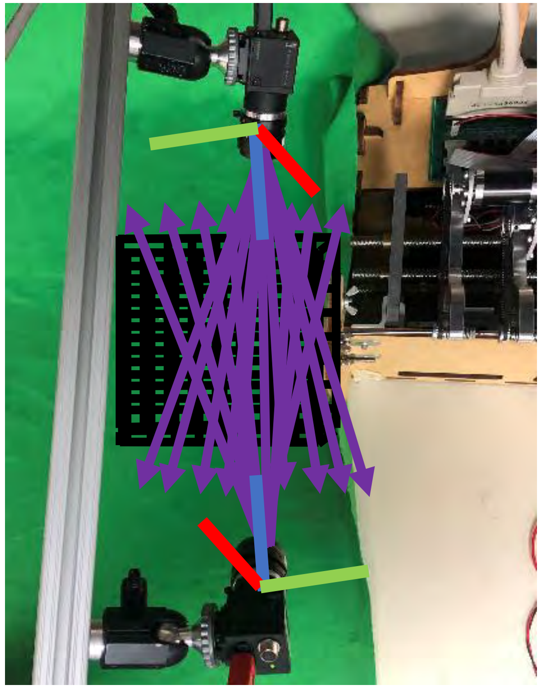
Training data: shape from silhouette



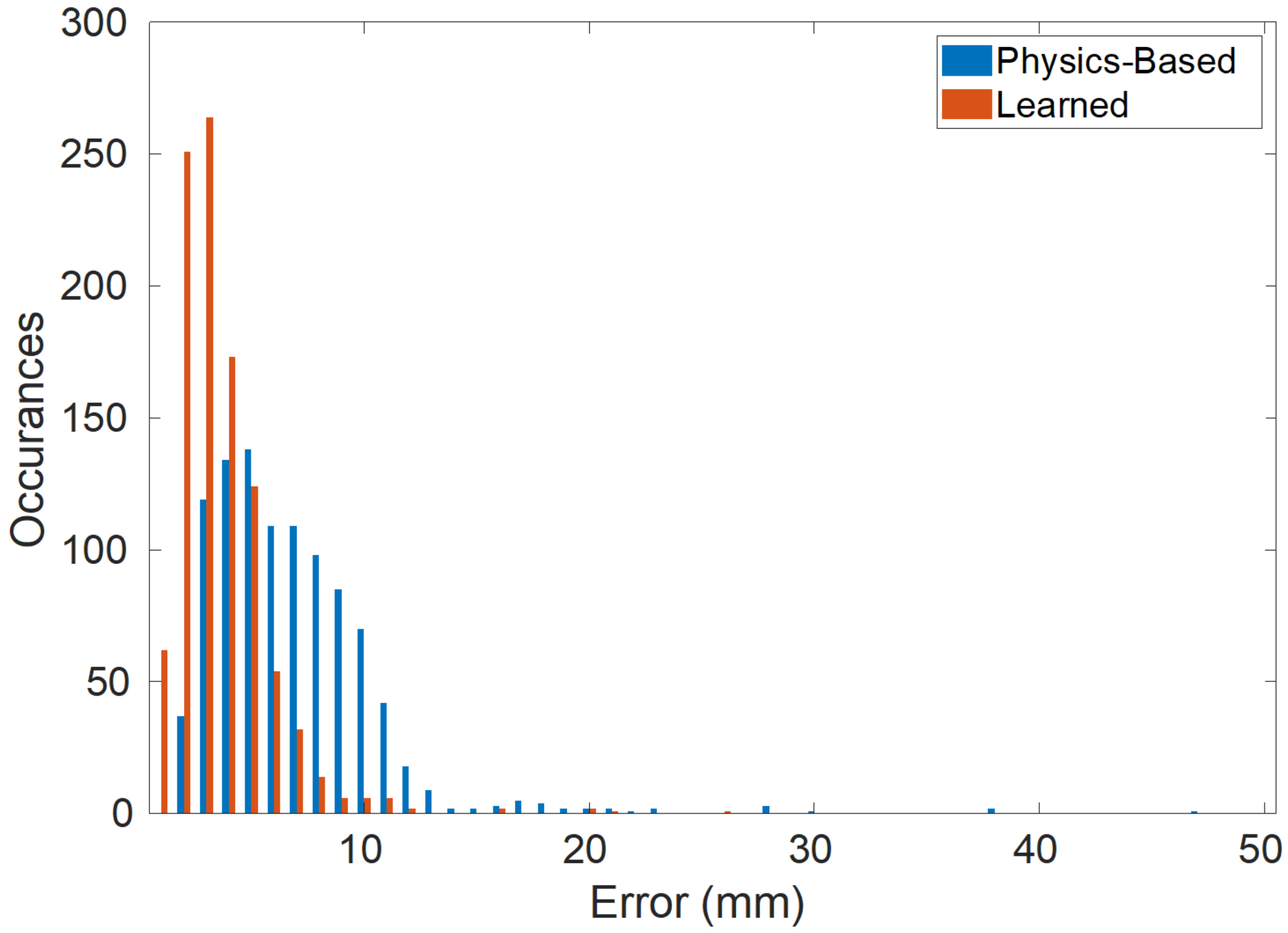


Training data: shape from silhouette

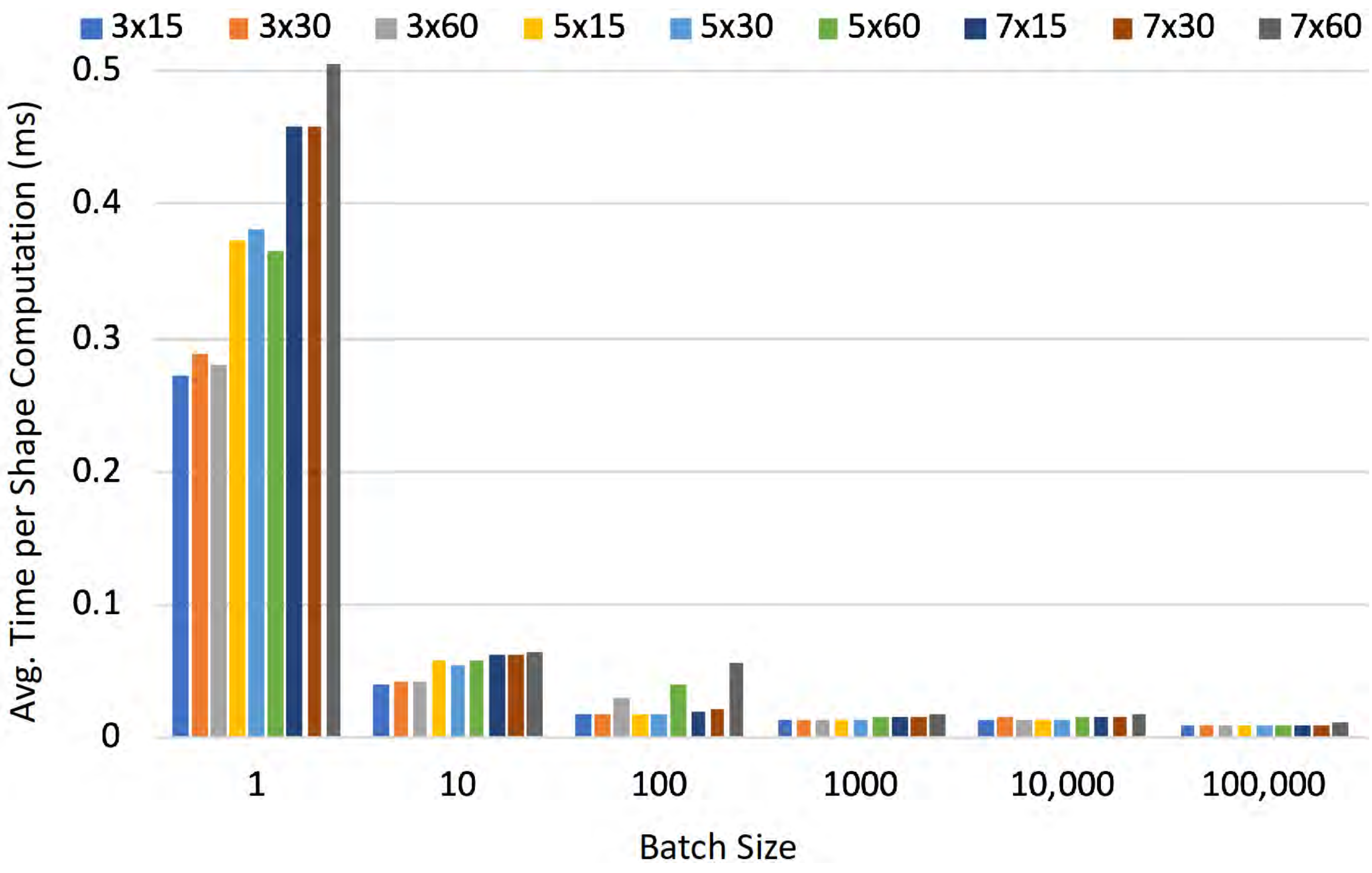




Learning a shape model



Learning a shape model



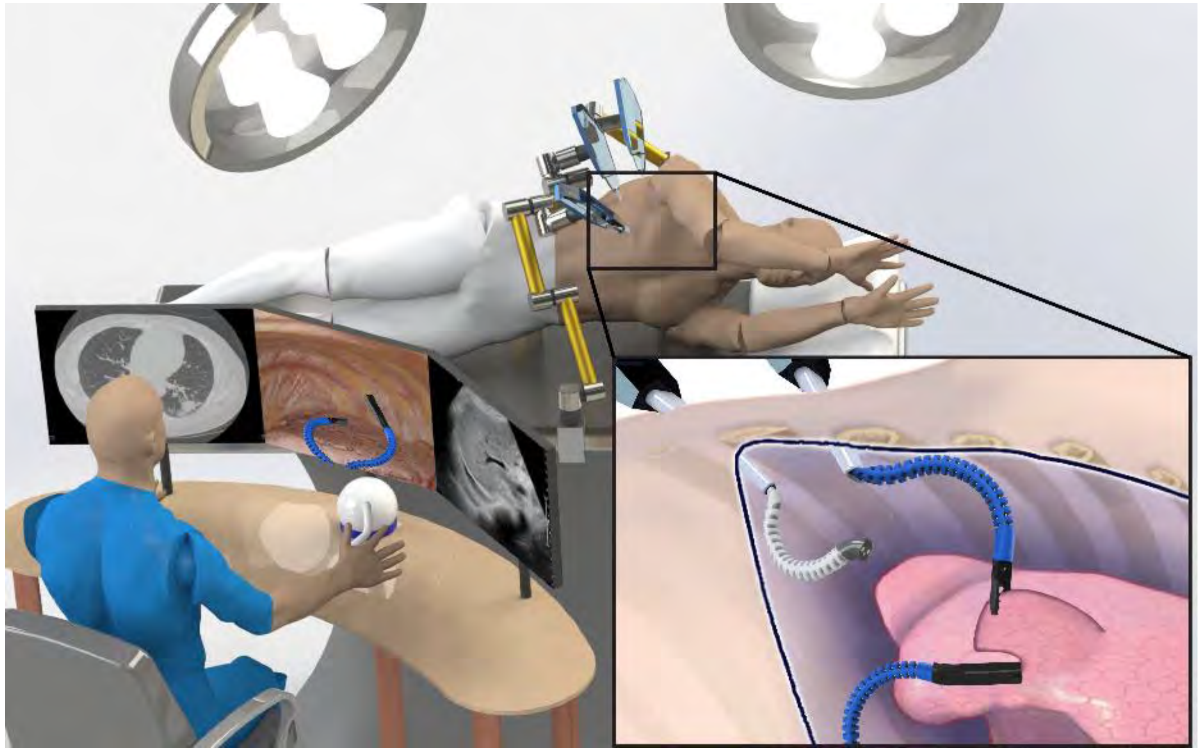
Physics-based model ~1.7ms

Subtask Automation Learned from Human Demonstrations

We can compute a motion, but *which* motion?

Full, reactive automated manipulation for complex, many-stepped tasks in uncertain environments is not yet feasible.

Leverage humans to demonstrate sub-task execution



Sensors

→

Human

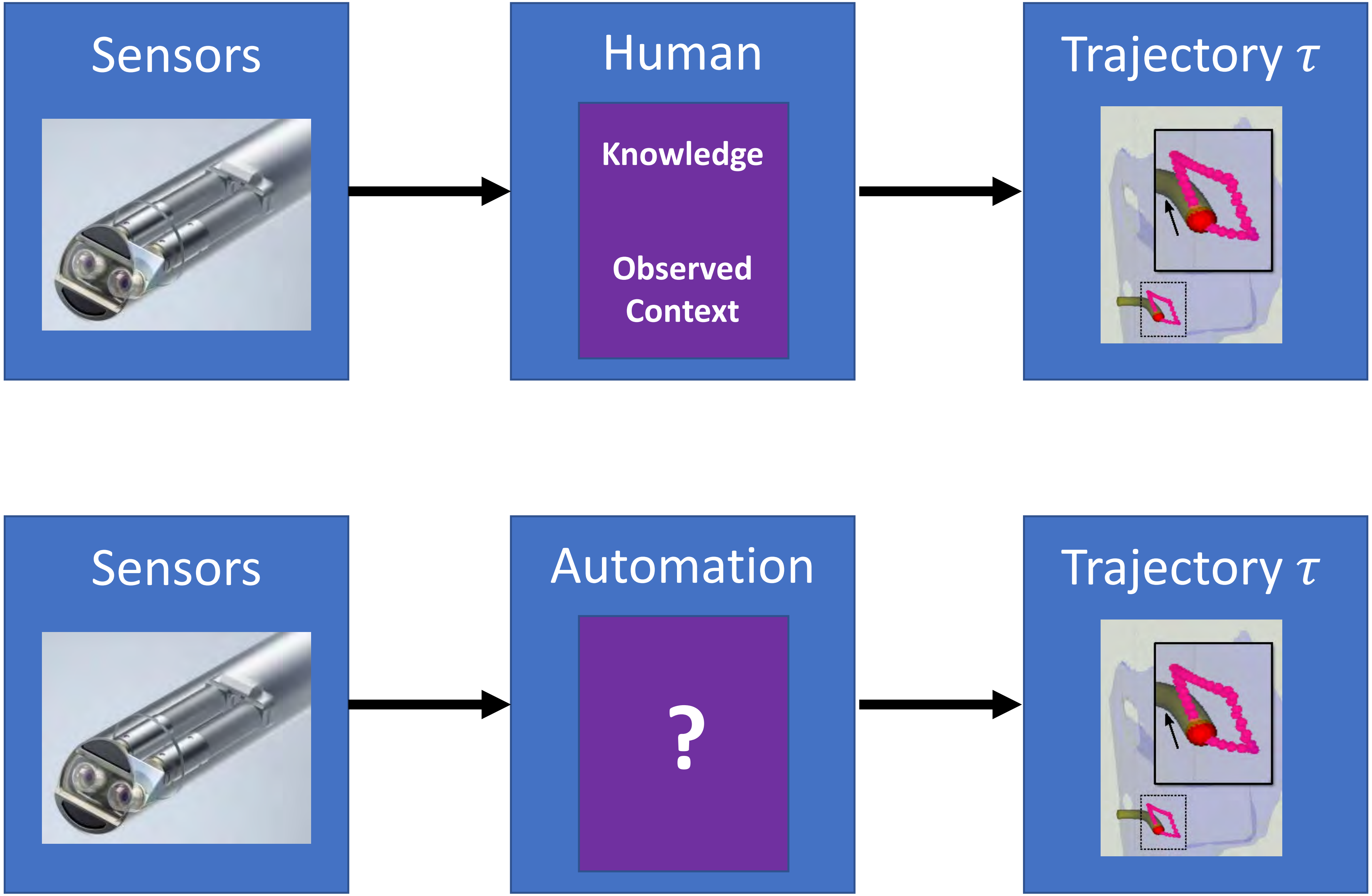
Knowledge

Observed Context

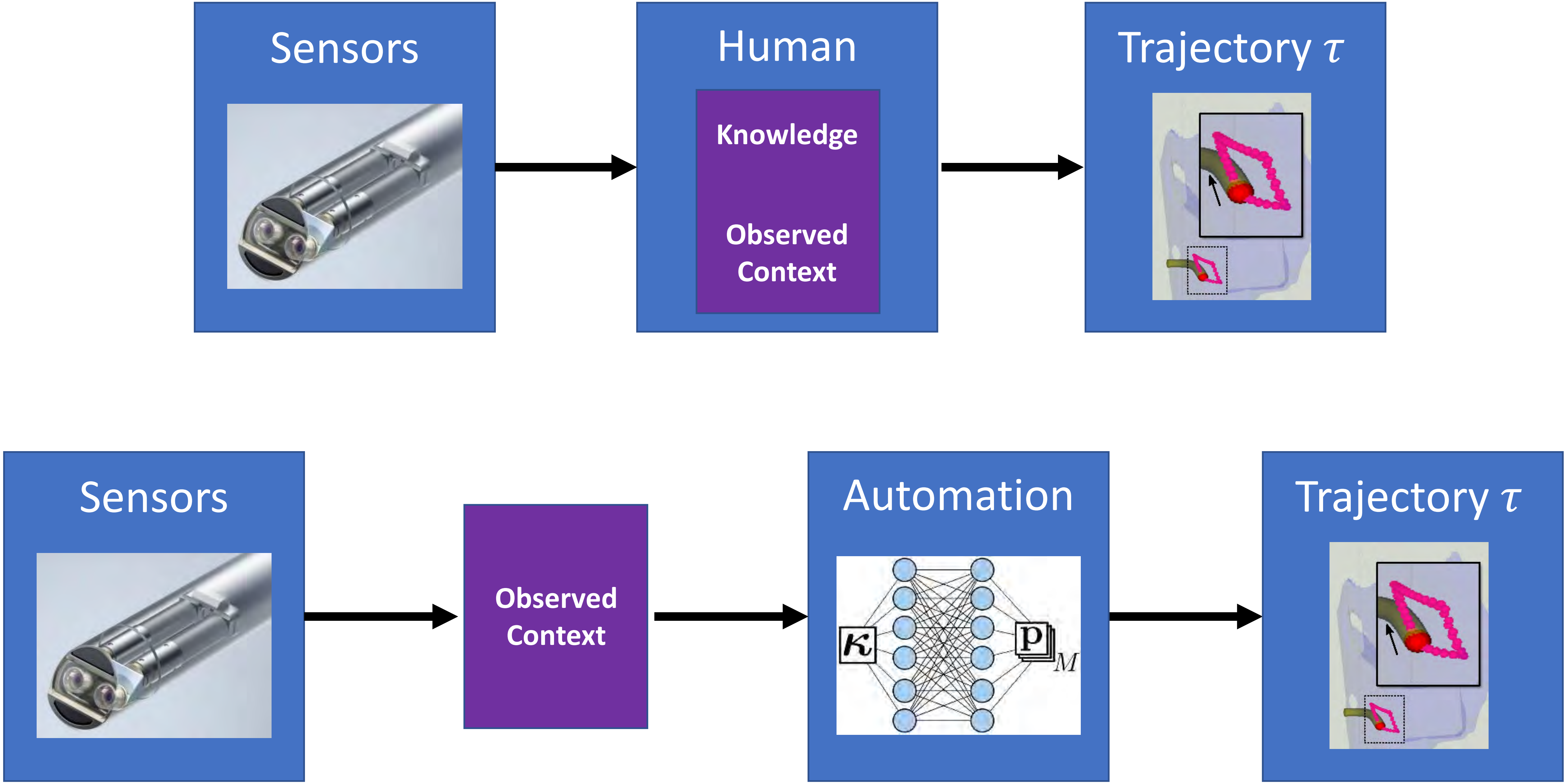
→

Trajectory τ

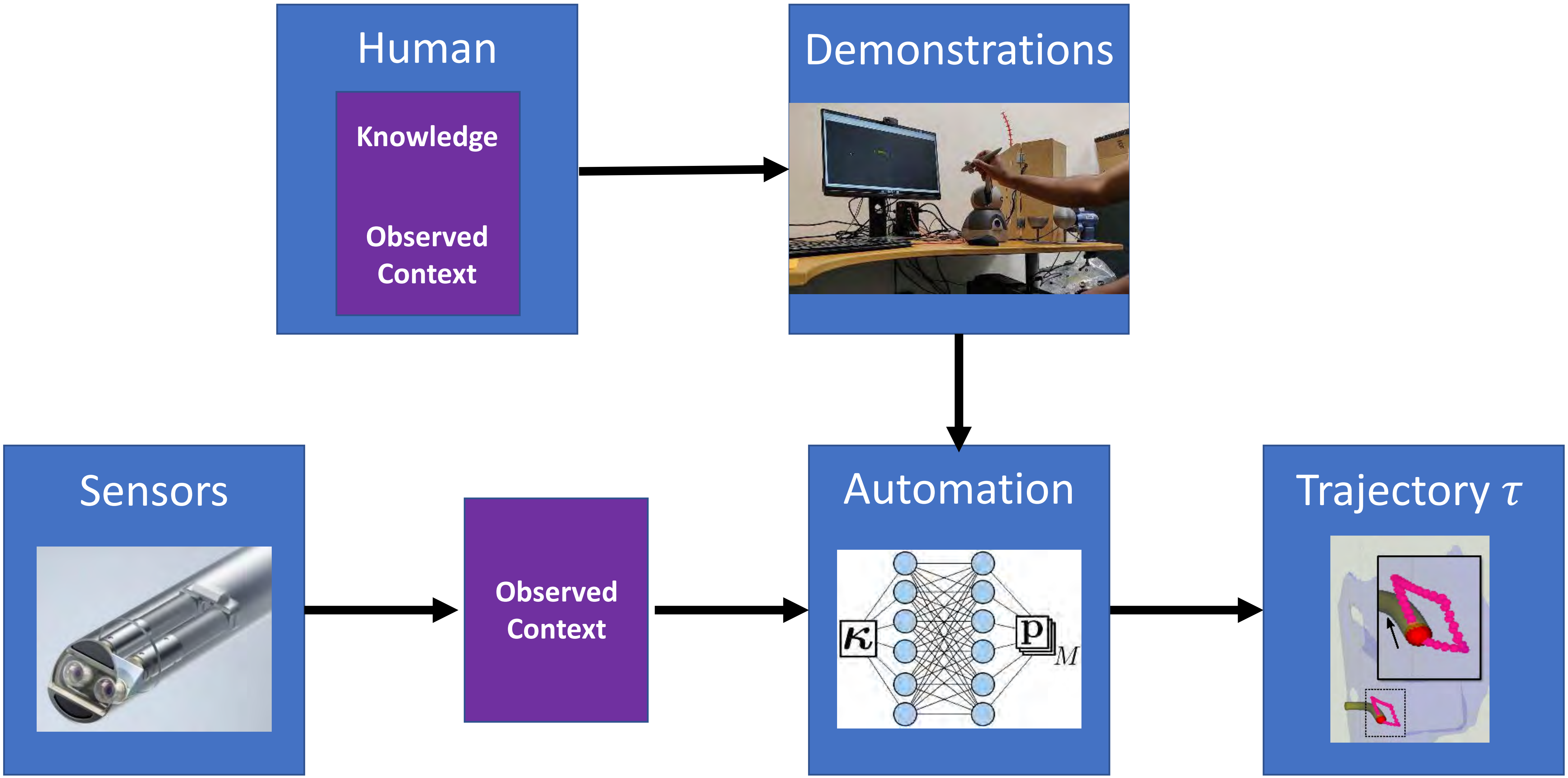
Endoscope image: https://www.intuitivesurgical.com/test-drive/pages/crystal-clear-3dhd.php

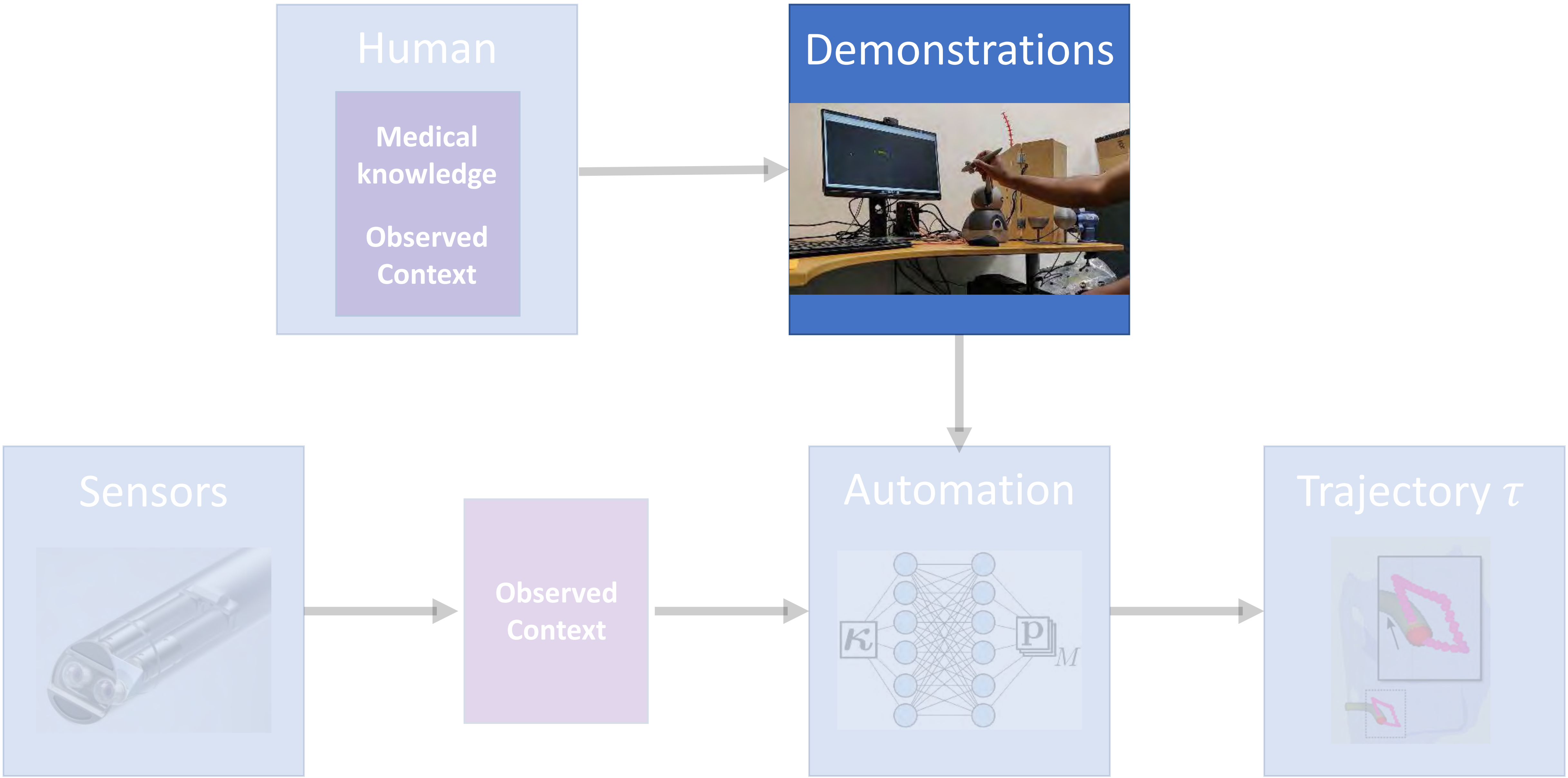


Sensors
Human
Knowledge
Observed Context
Trajectory τ
Sensors
Automation
?
Trajectory τ



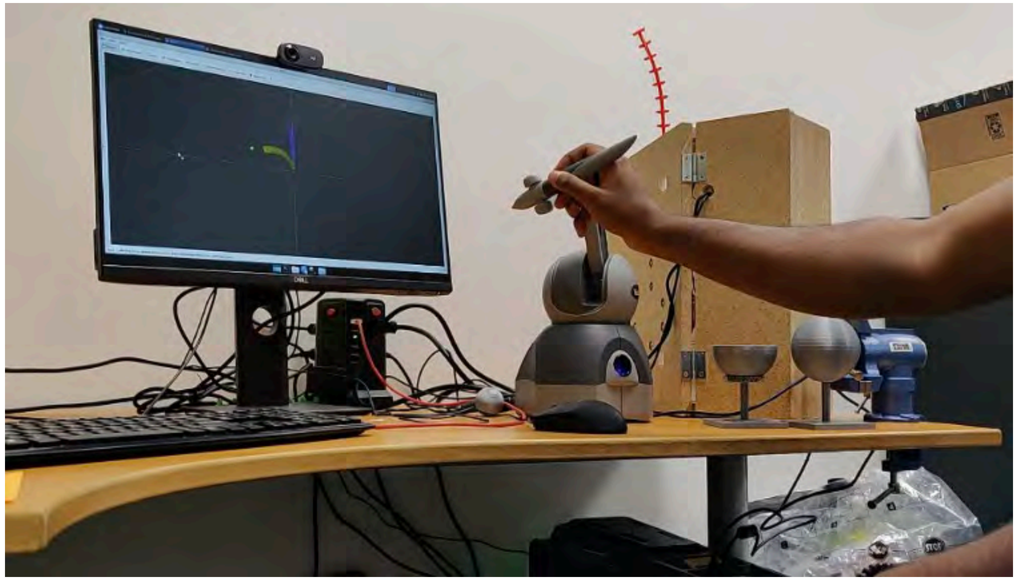
Sensors
Human
Knowledge
Observed Context
Trajectory τ
Sensors
Observed Context
Automation
$\boldsymbol{\kappa}$
$\mathbf{p}_M$
Trajectory τ


Human
Knowledge
Observed Context
Demonstrations
Sensors
Observed Context
Automation
$\boldsymbol{\kappa}$
$\mathbf{p}_M$
Trajectory τ

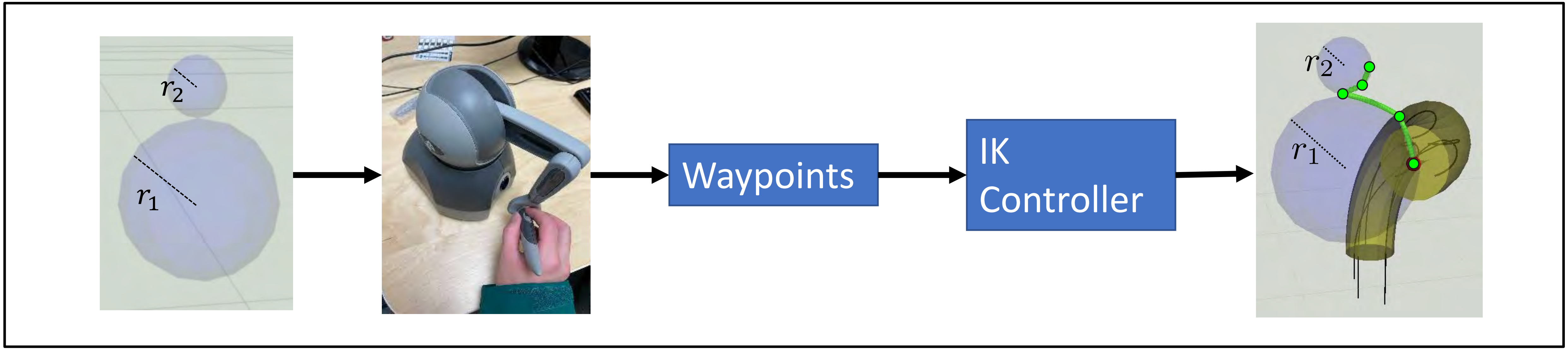


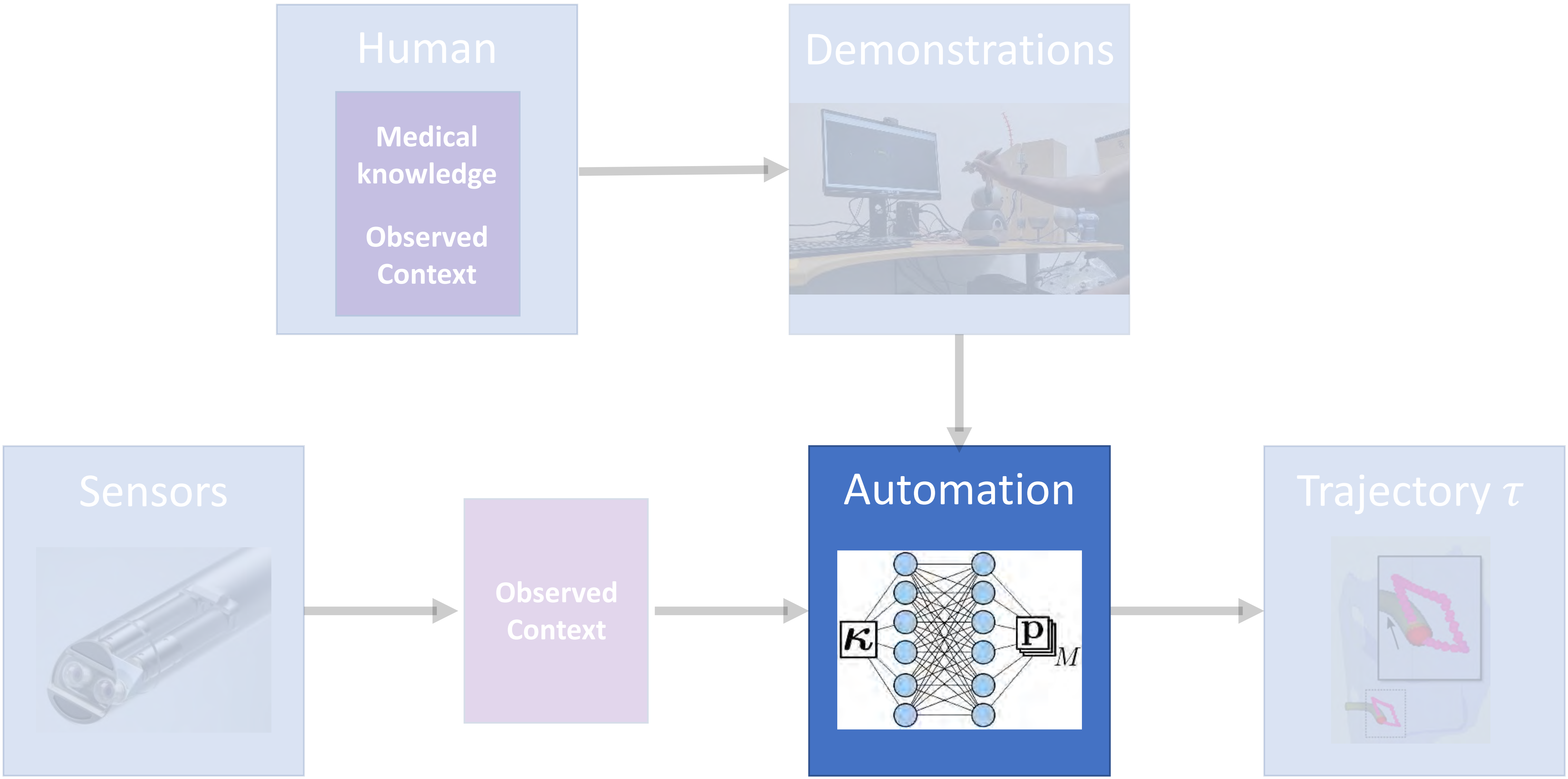
Human
Medical knowledge
Observed Context
Demonstrations
Sensors
Observed Context
Automation
κ
p_M
Trajectory τ

Human demonstration collection



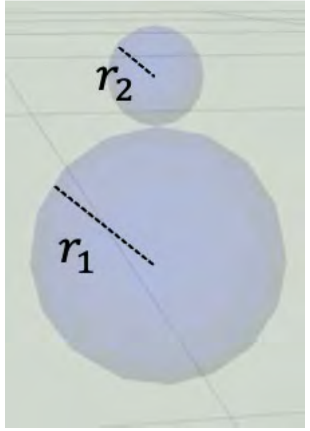
Demonstrations with different contexts



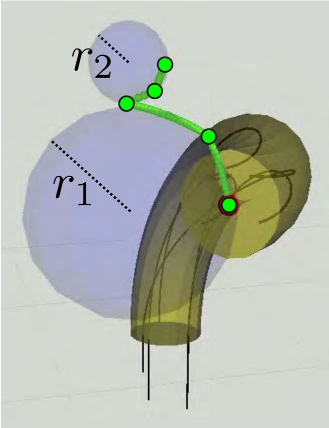


Human
Medical knowledge
Observed Context
Demonstrations
Sensors
Observed Context
Automation
κ
$\mathbf{p}_M$
Trajectory τ

$$f(\kappa) \rightarrow T' \quad \kappa \triangleq context$$



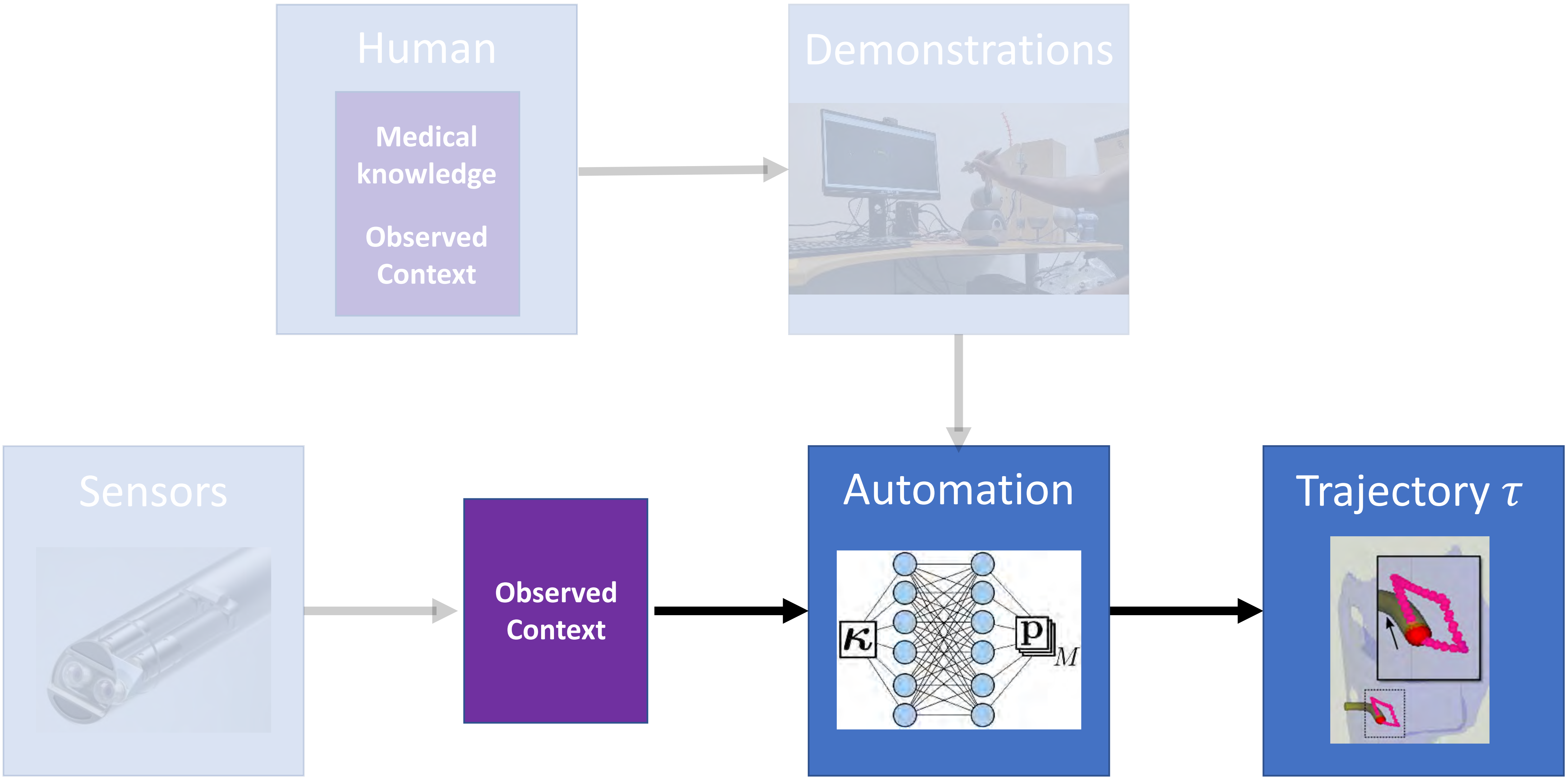
$$T' \triangleq trajectory$$



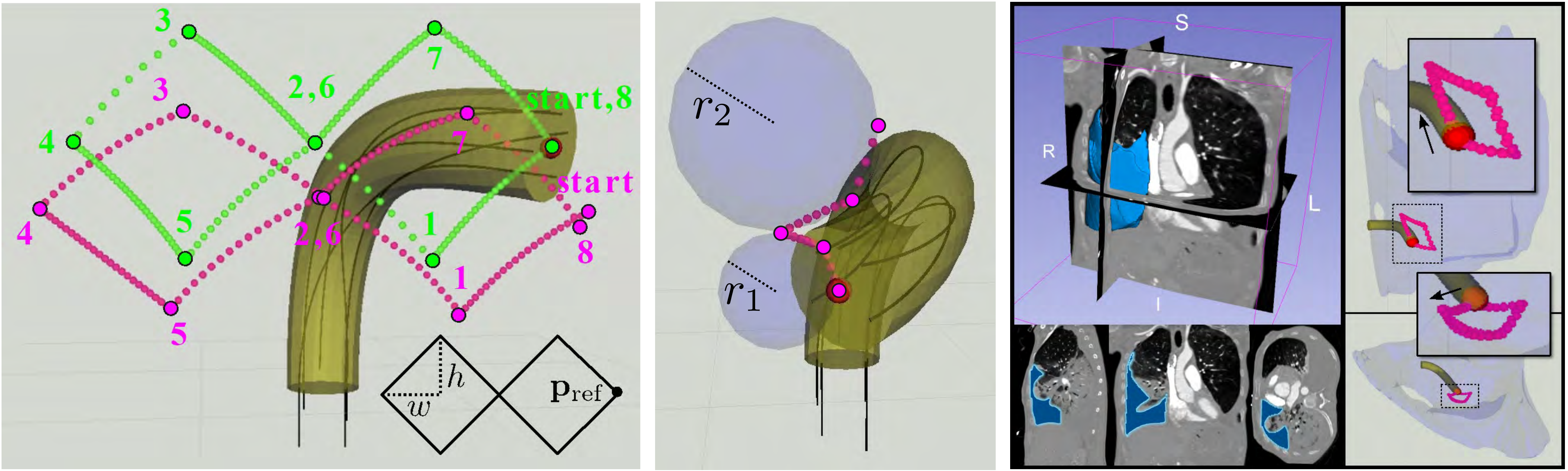
$$\mathrm{argmin}_W \sum_{i=1}^{D} \|f(\kappa) - T_i\|_2^2 + \alpha\|W\|_2^2$$

Predictor f:

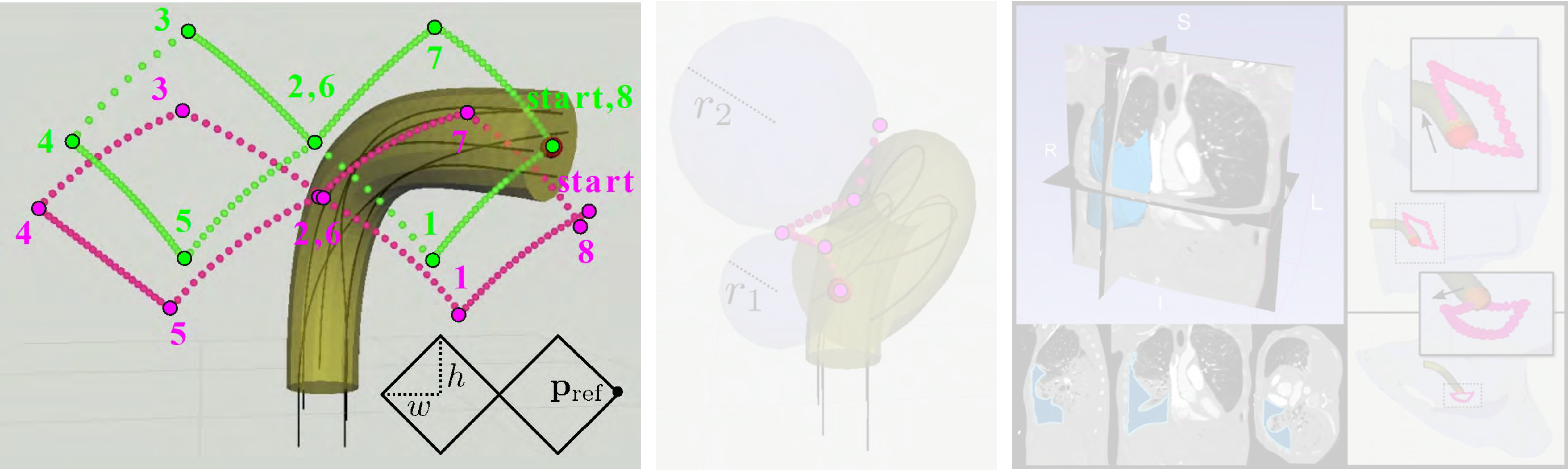
- Linear Ridge Regression
- Radial-Basis Function Kernel Ridge Regression (RBF)
- Neural Trajectory Network



Human
Medical knowledge
Observed Context
Demonstrations
Sensors
Observed Context
Automation
$\boldsymbol{\kappa}$
$\mathbf{p}_M$
Trajectory τ

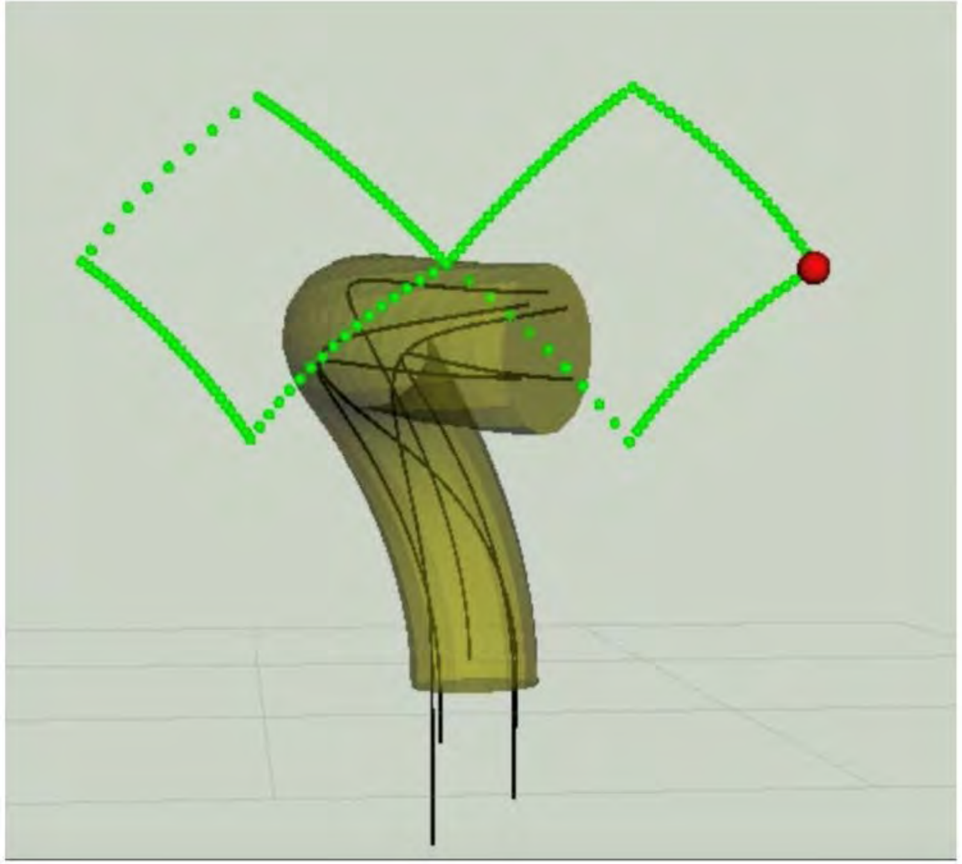


Steps toward more complex tasks
Environmental geometry dependent

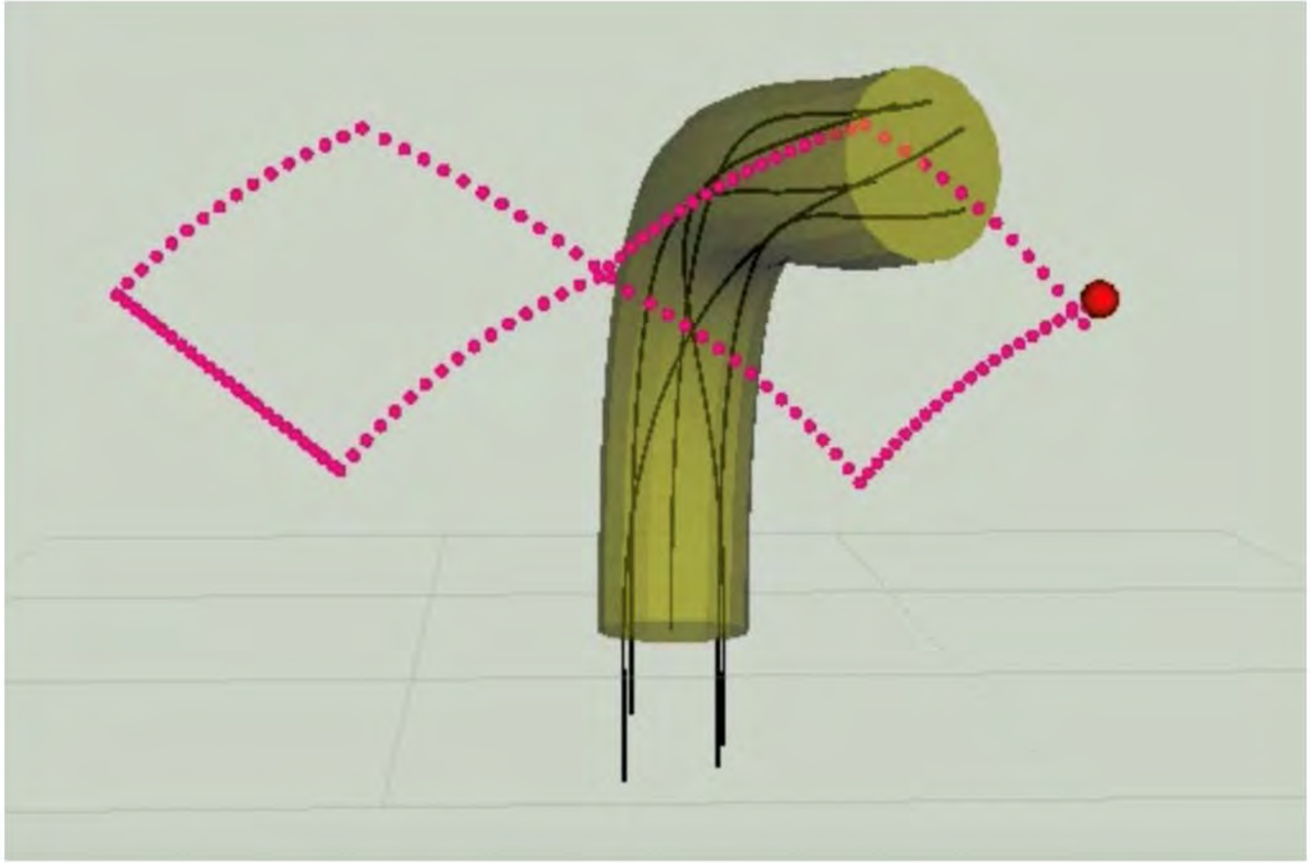


Context:

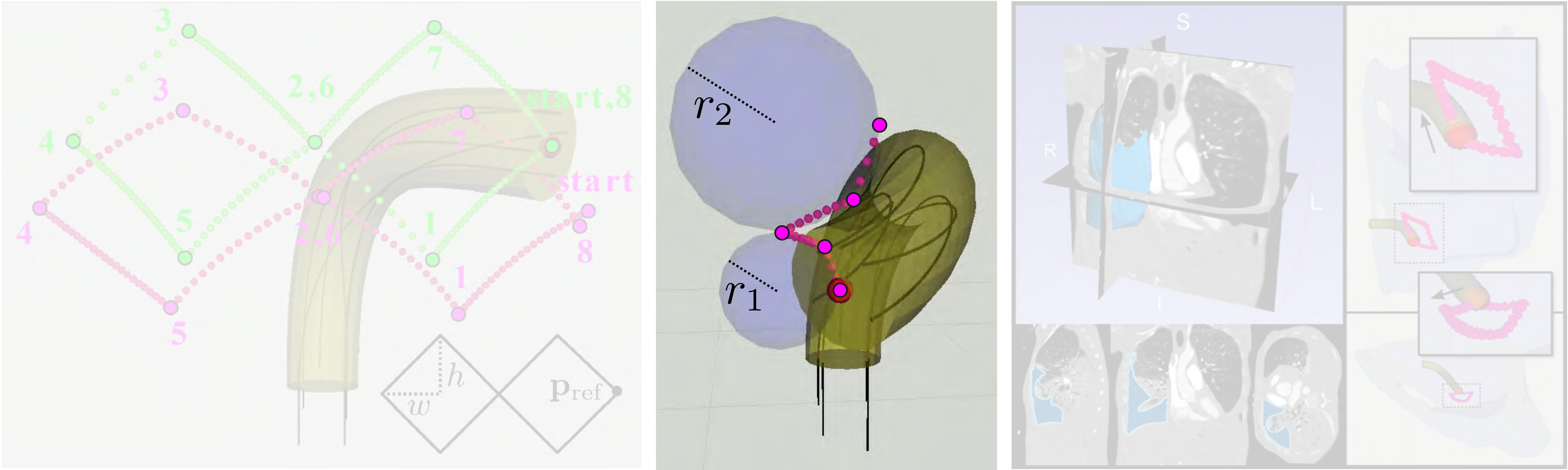
Width, height, and reference point



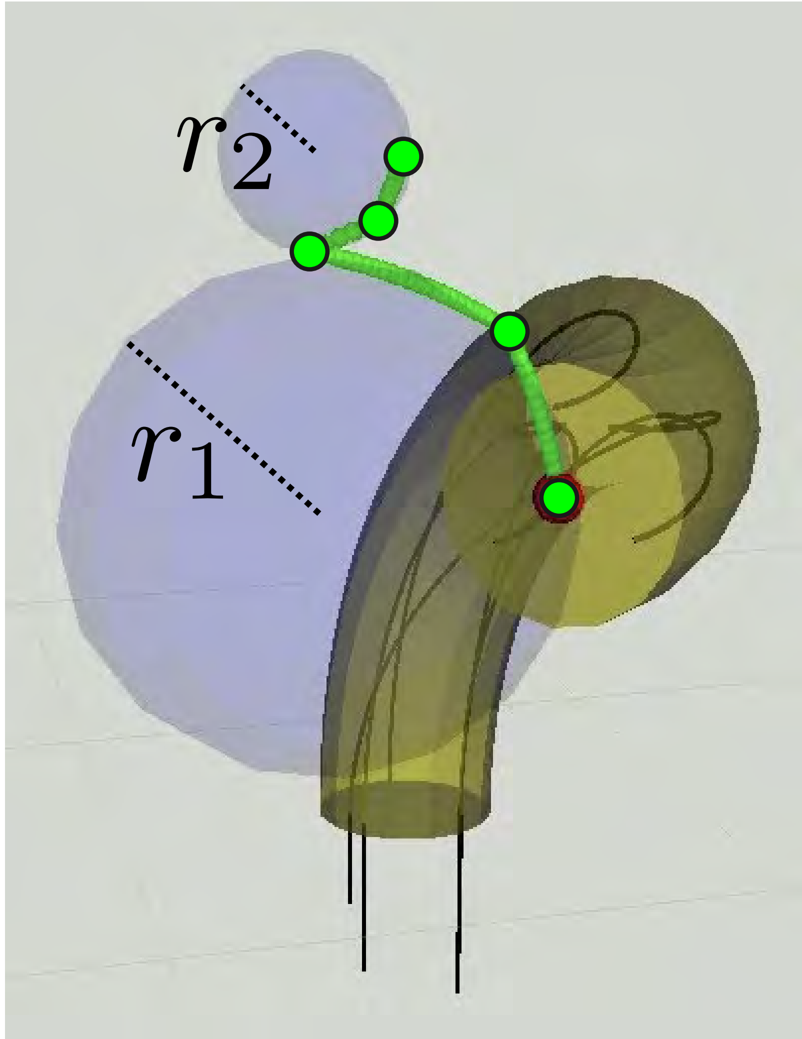
Demonstration



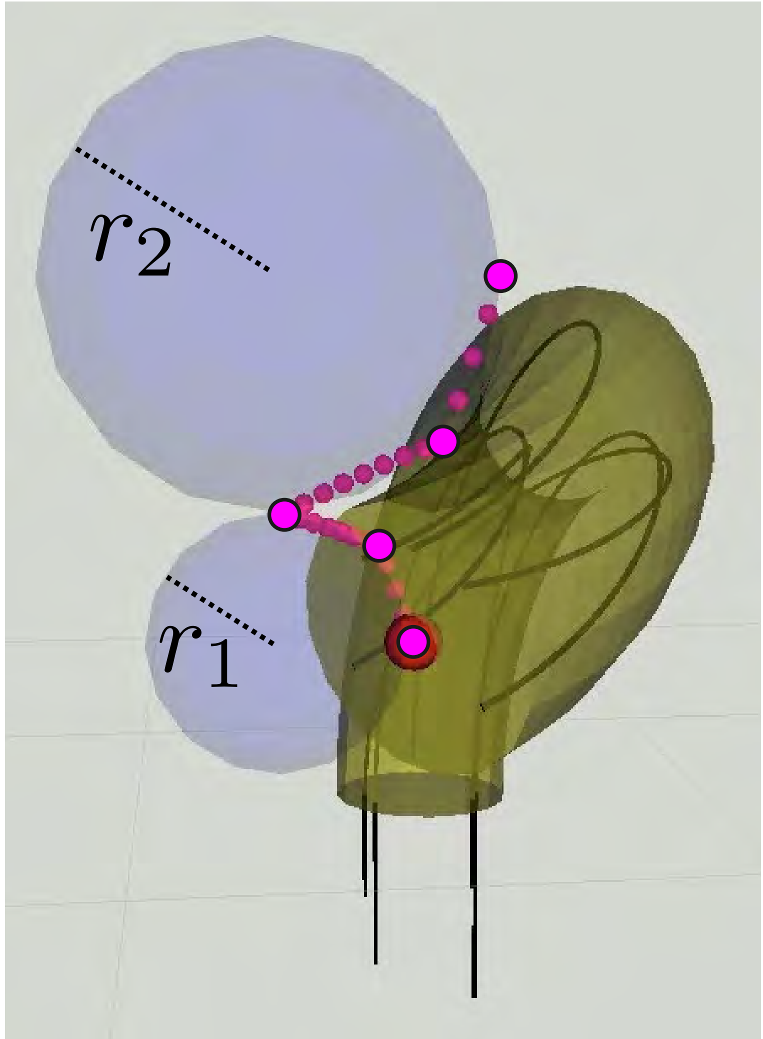
Trajectory network



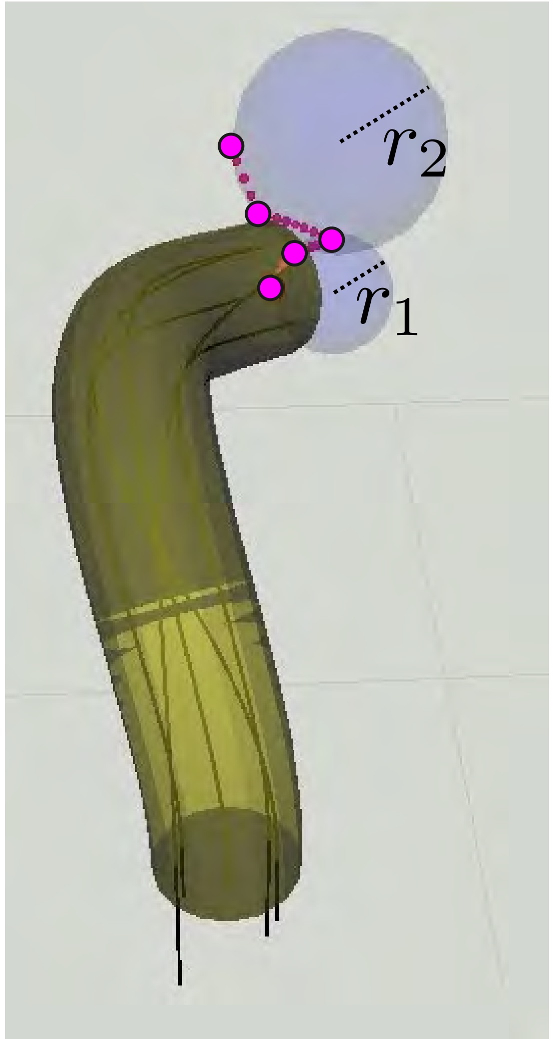
Context:
Reference point
Radii of the two spheres



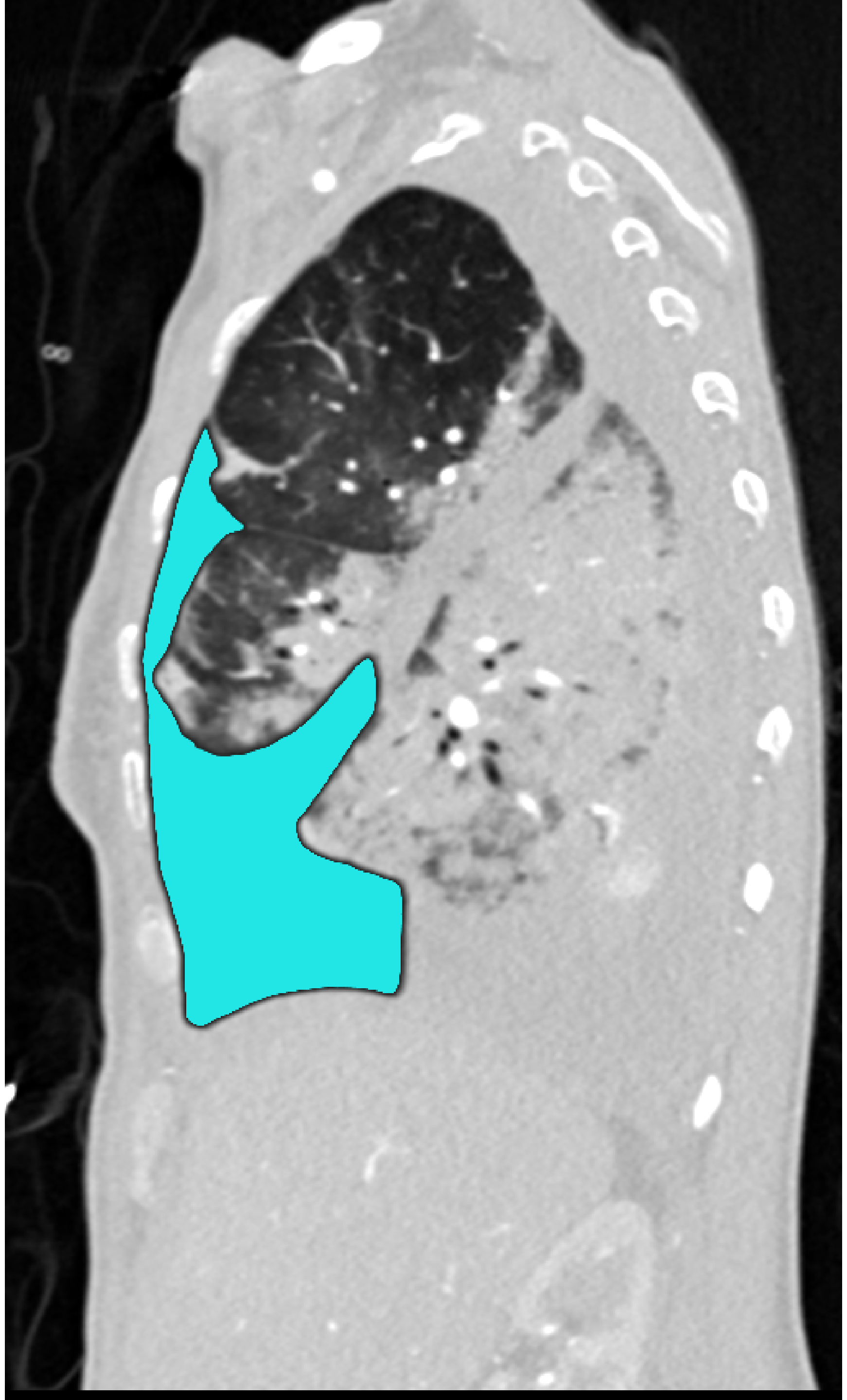
Demonstration

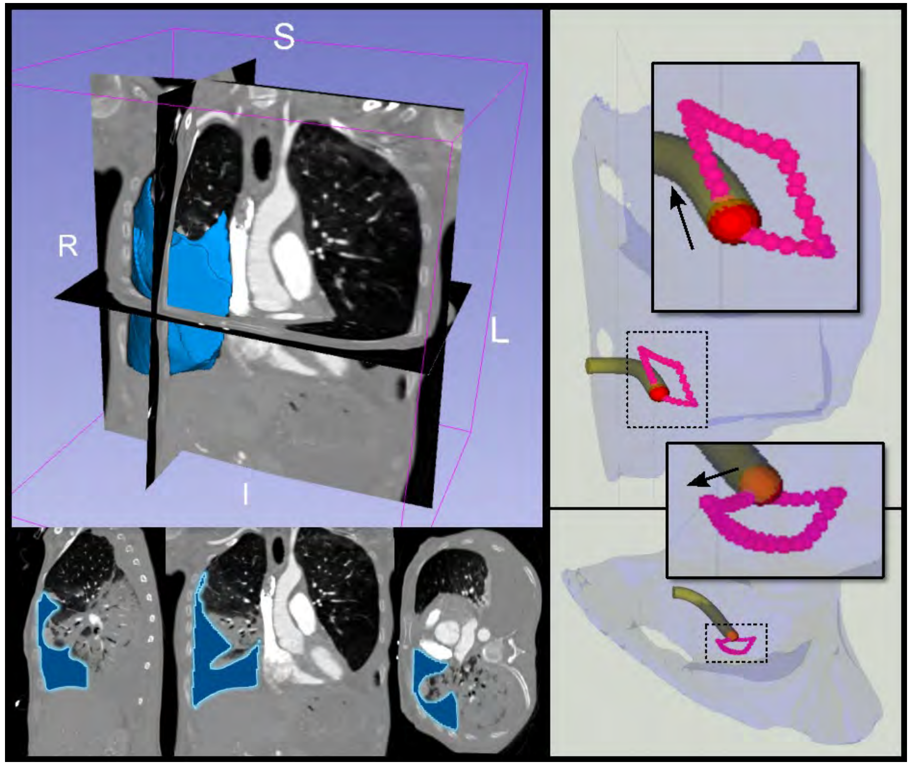


Trajectory network



Trajectory network

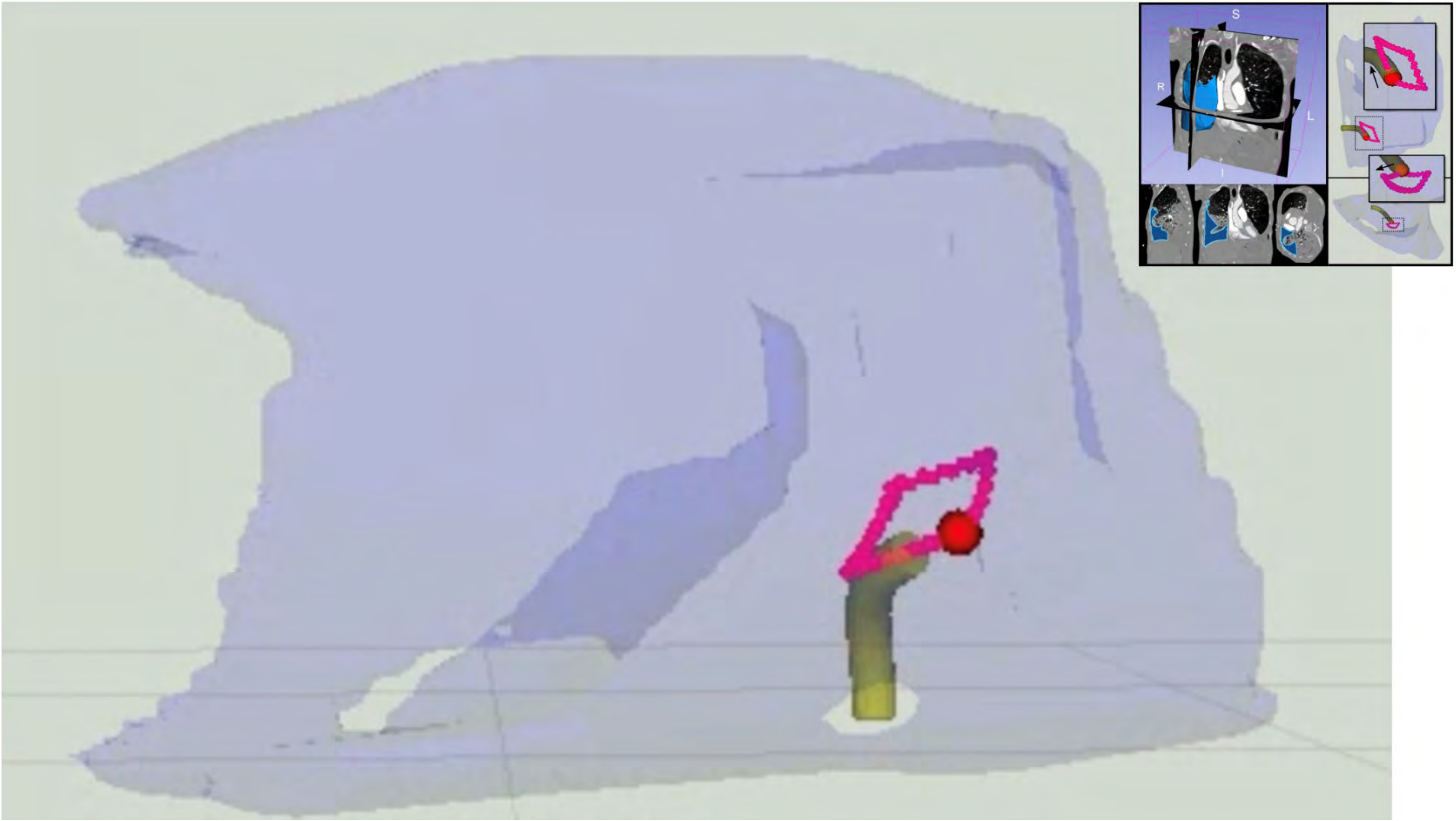



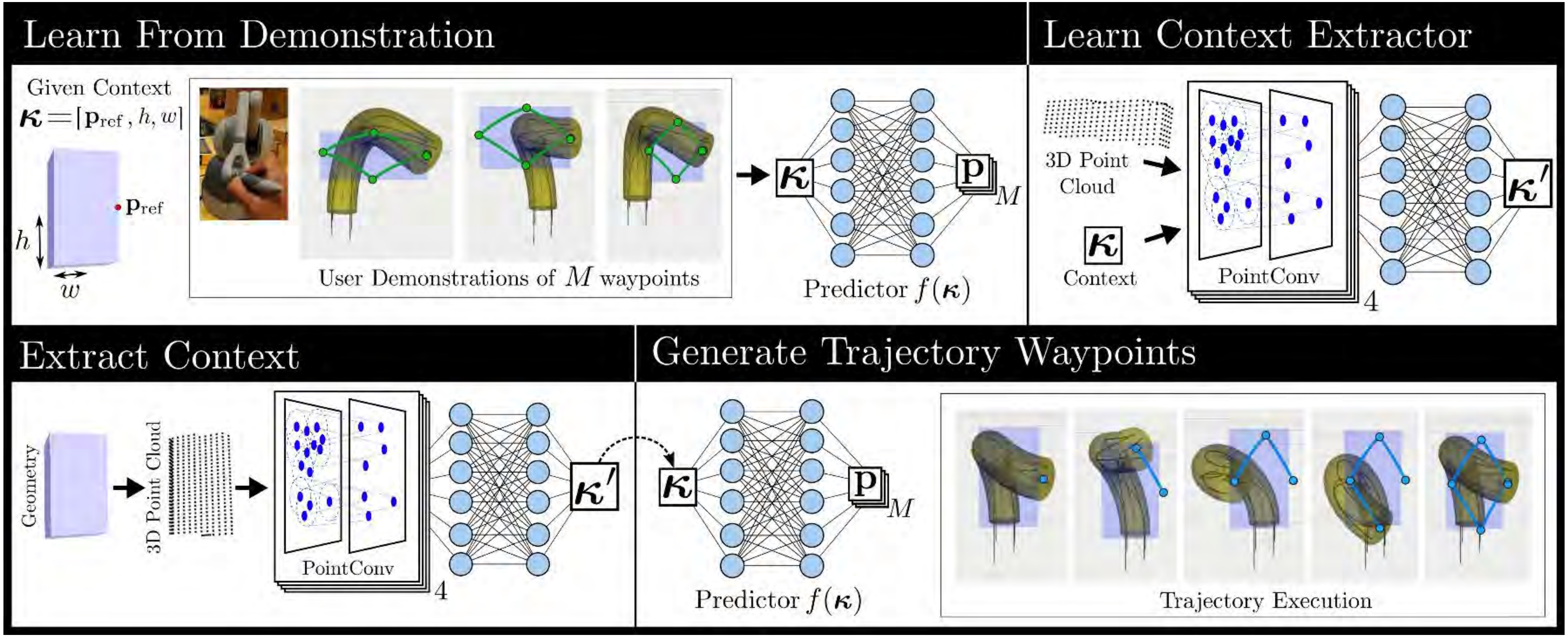


Context:

Reference point

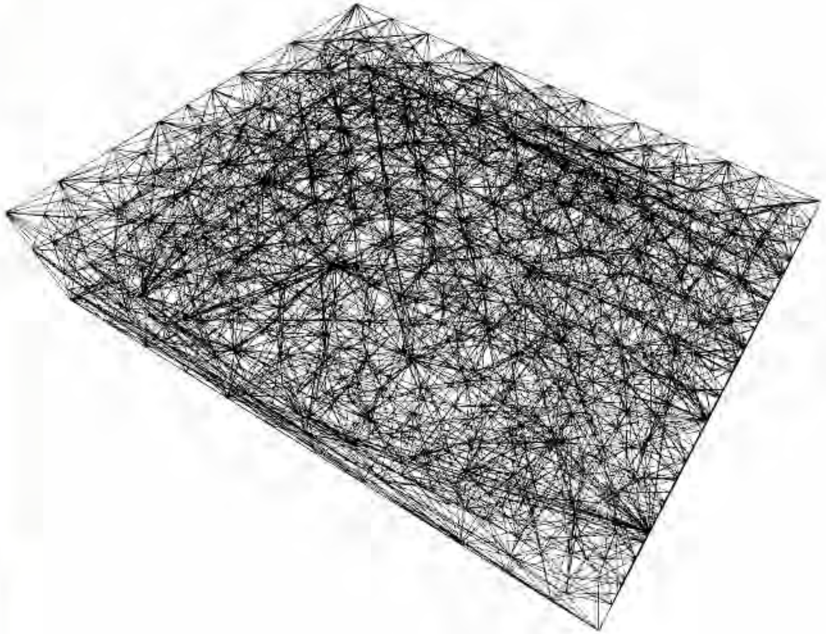
Scale of the anatomy

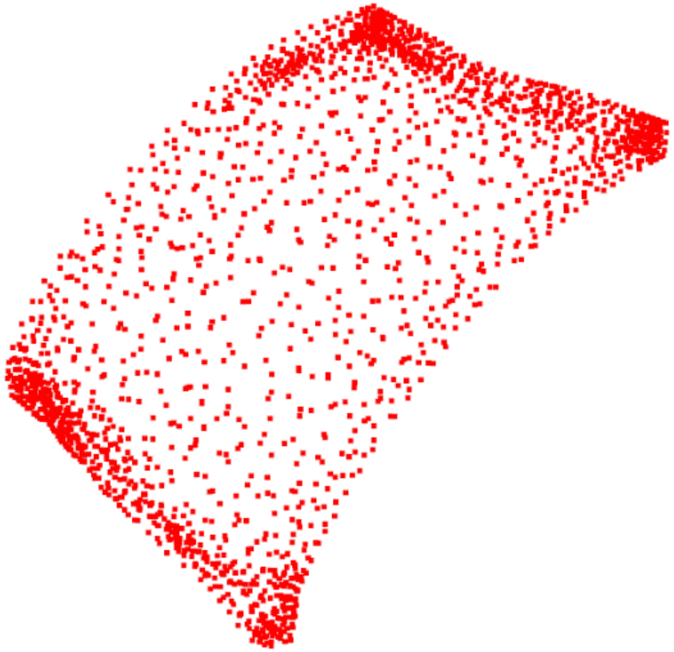

S
R
L
I



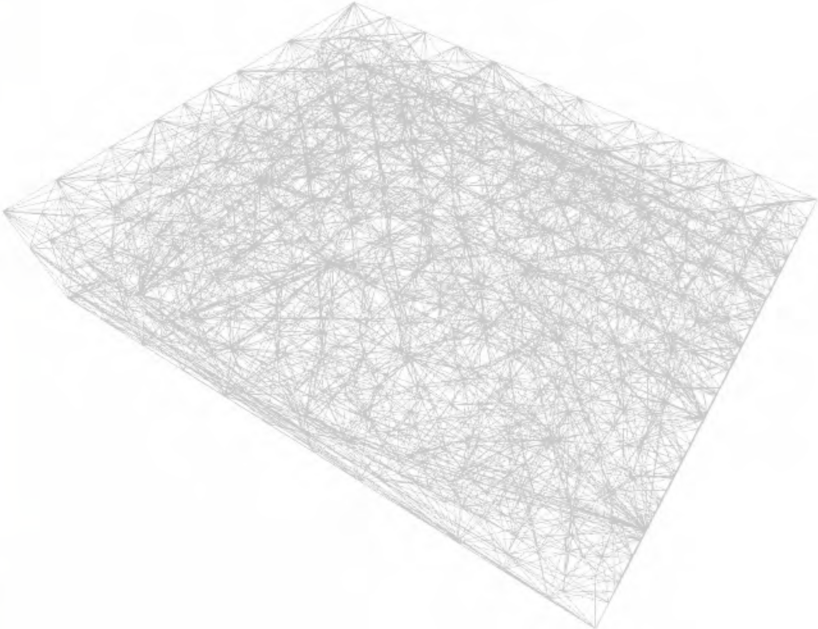
Learn From Demonstration
Given Context
$\boldsymbol{\kappa} = [\mathbf{p}_{\mathrm{ref}}, h, w]$
$\mathbf{p}_{\mathrm{ref}}$
h
w
User Demonstrations of M waypoints
$\boldsymbol{\kappa}$
$\mathbf{p}$
M
Predictor $f(\boldsymbol{\kappa})$
Learn Context Extractor
3D Point Cloud
$\boldsymbol{\kappa}$
Context
PointConv
4
$\boldsymbol{\kappa}'$
Extract Context
Geometry
3D Point Cloud
PointConv
4
$\boldsymbol{\kappa}'$
Generate Trajectory Waypoints
$\boldsymbol{\kappa}$
$\mathbf{p}$
M
Predictor $f(\boldsymbol{\kappa})$
Trajectory Execution

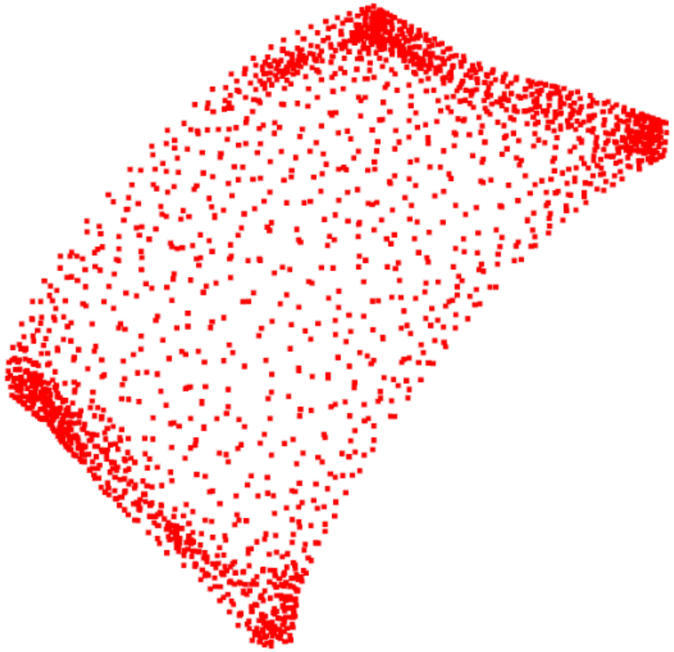
Representing deformable objects



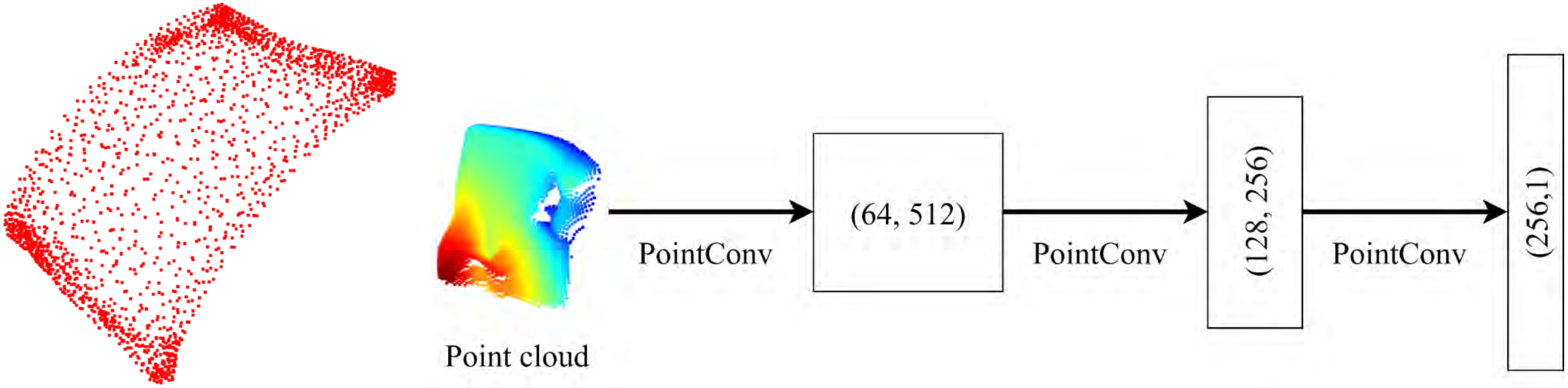


Representing deformable objects

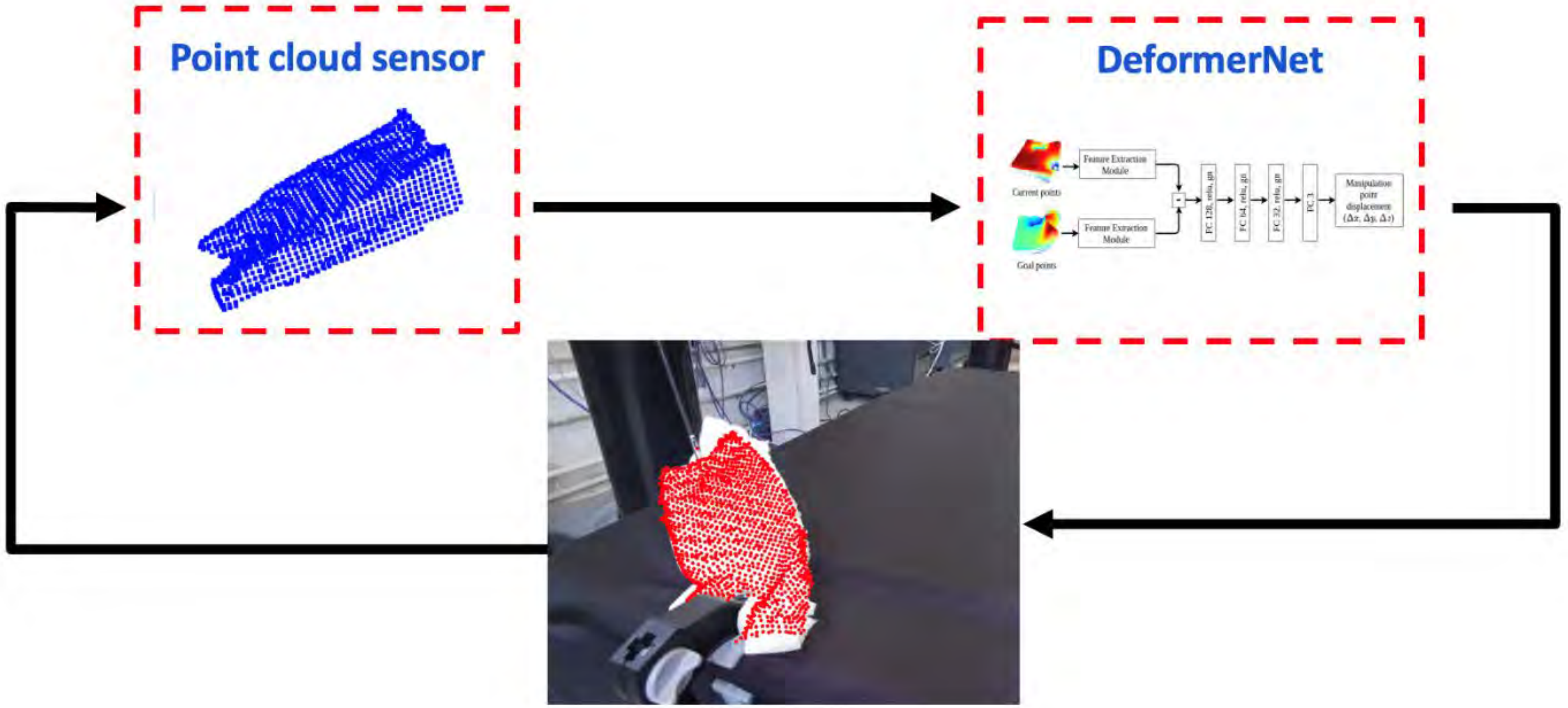




Low-dimensional point cloud representation



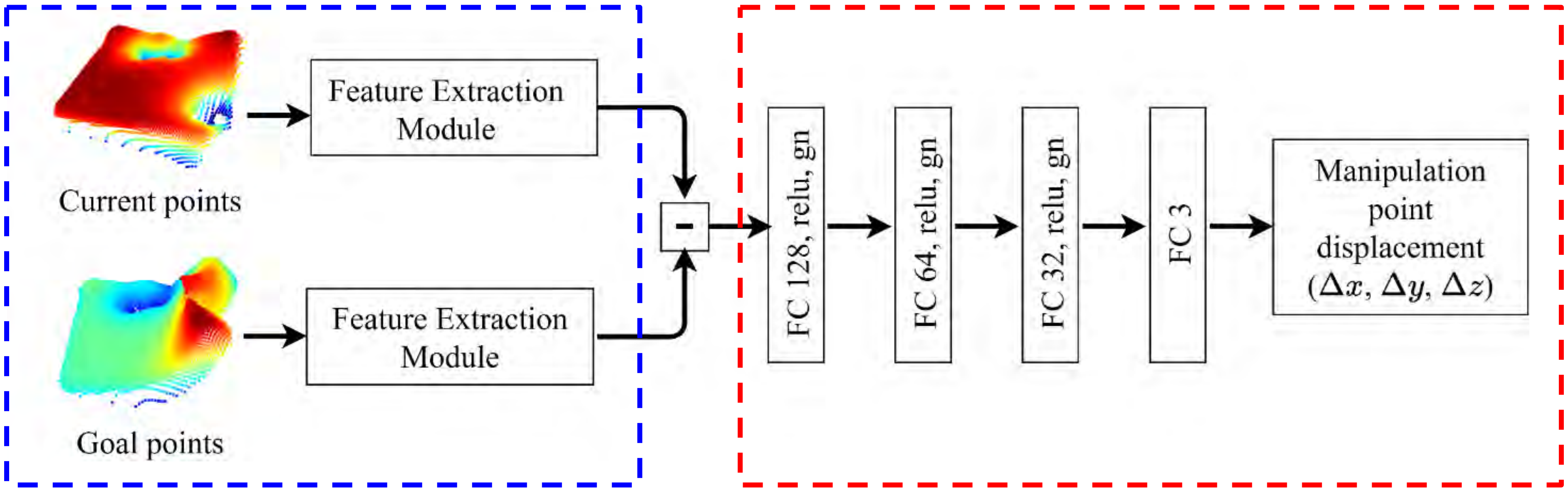
Shape servo control loop



DeformerNet

Architecture of *DeformerNet*

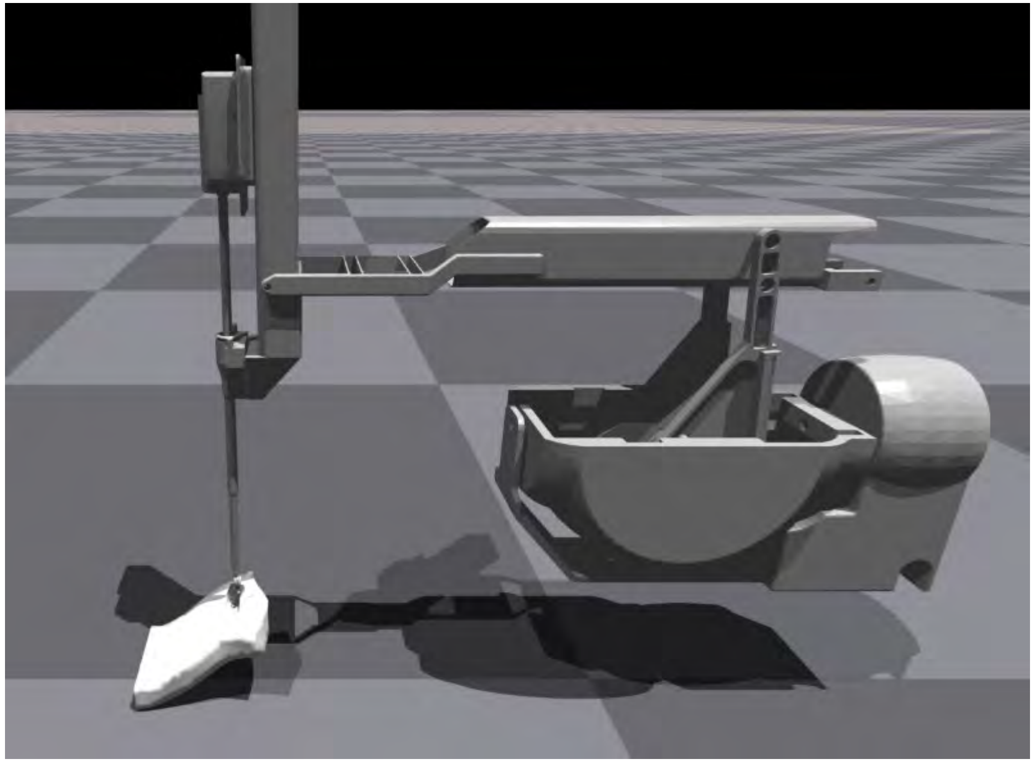
Feature Extraction



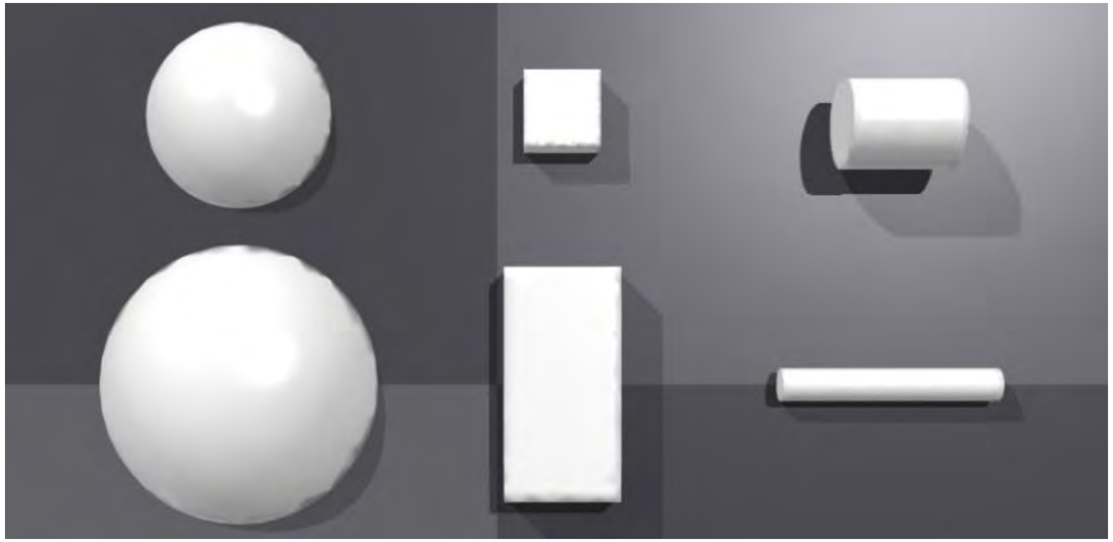
Deformation Control

Training data collection

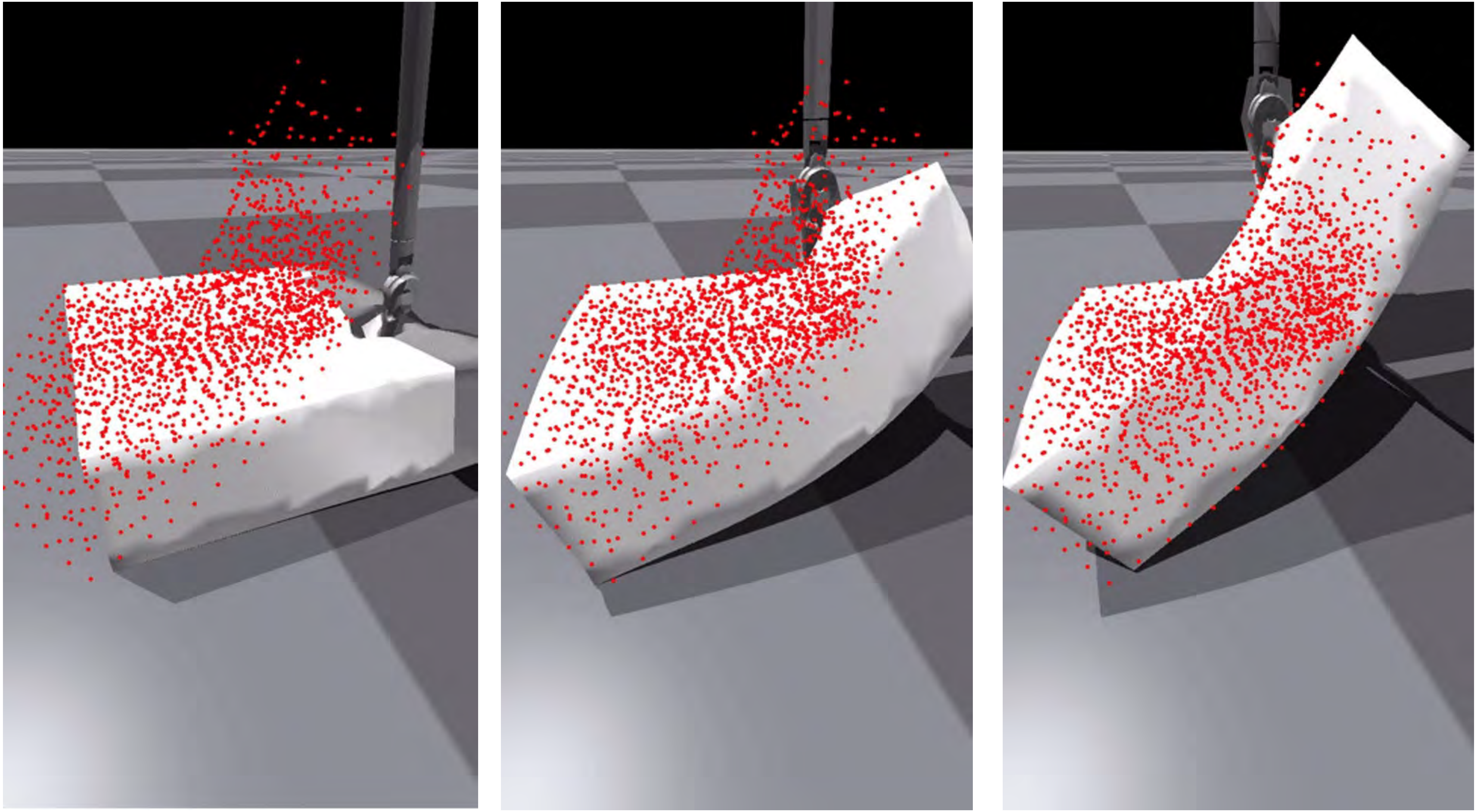
- Leverage large-scale deformable object simulation to generate training data

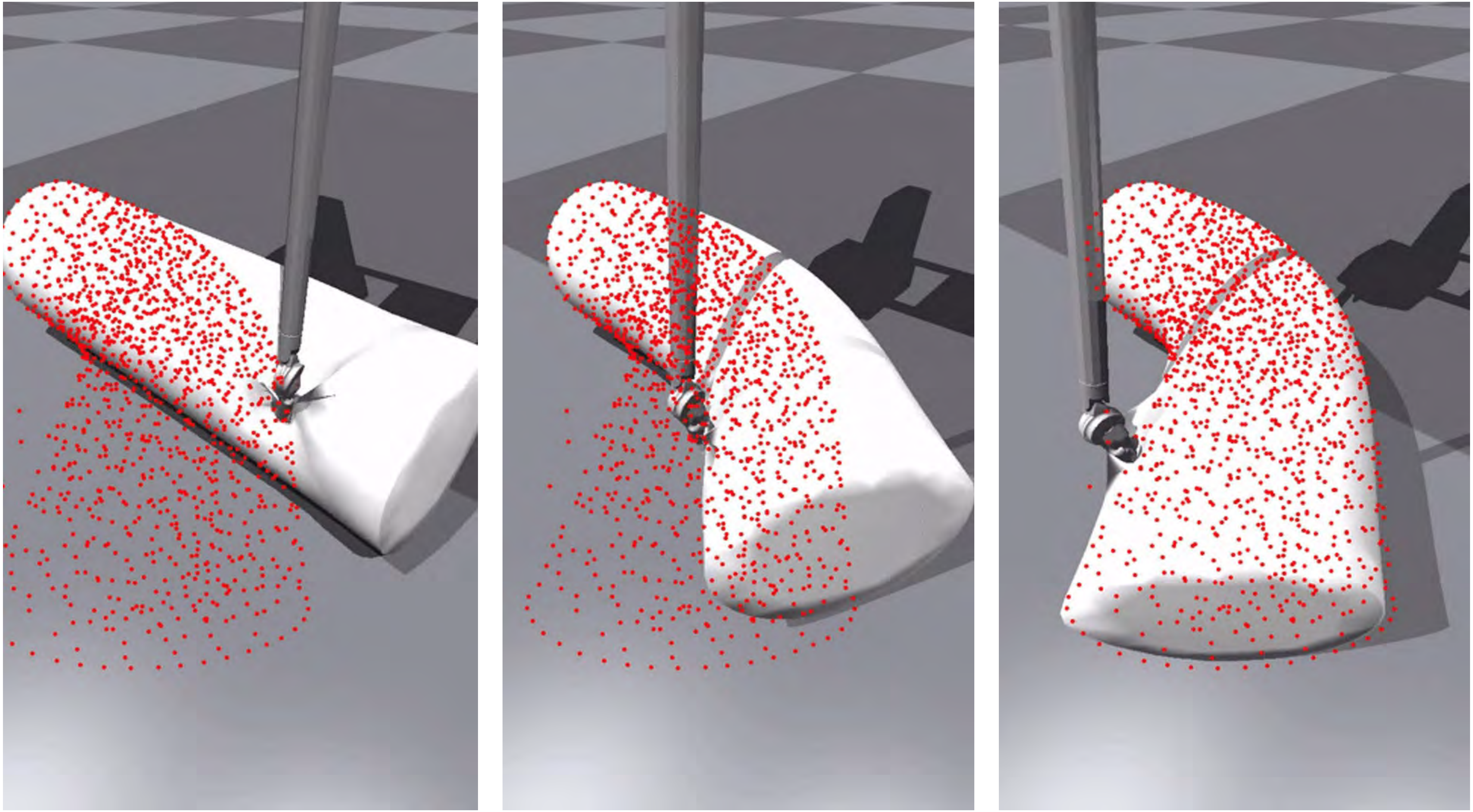


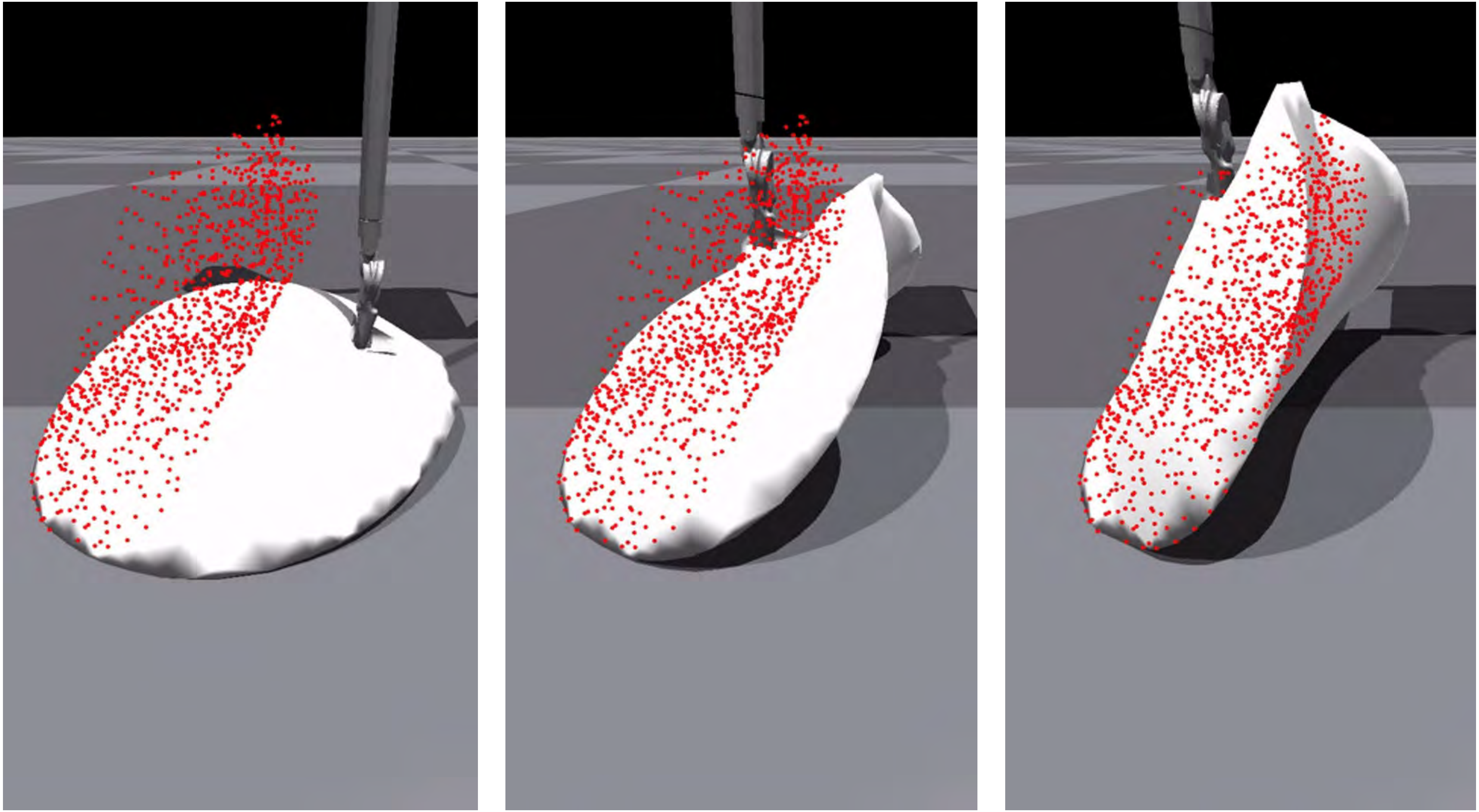
daVinci Research Kit (dVRK) surgical robot

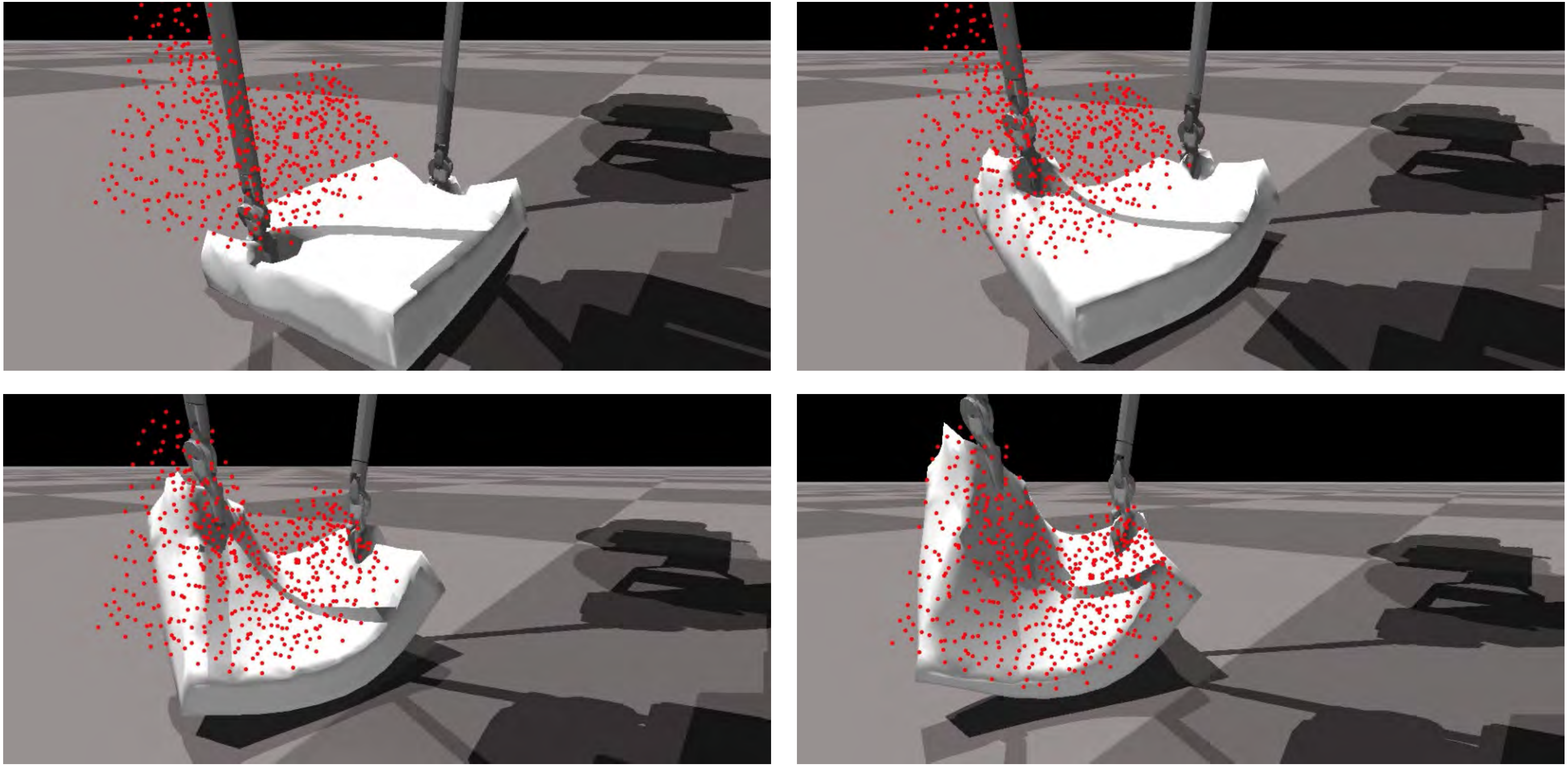


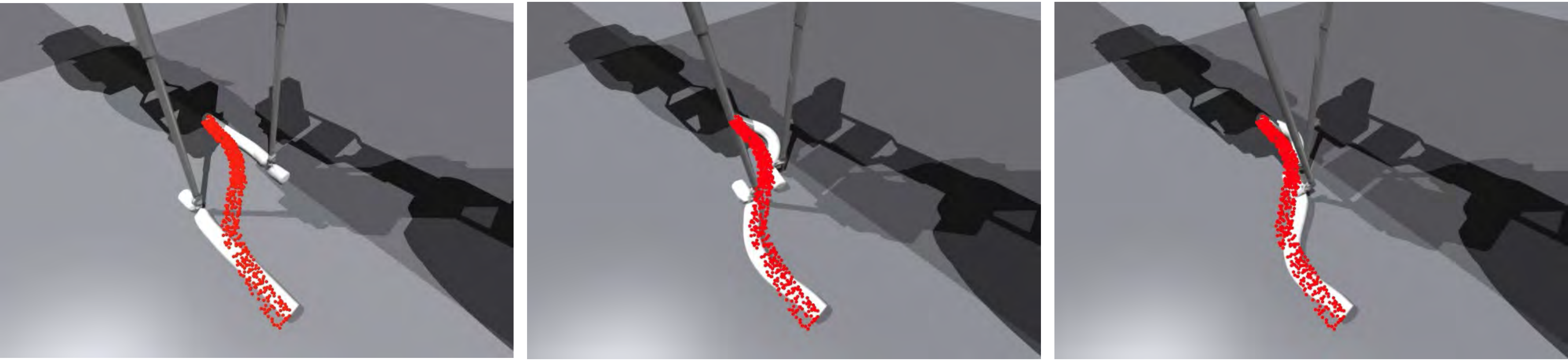
DeformerNet is trained on a variety of object geometries and stiffness

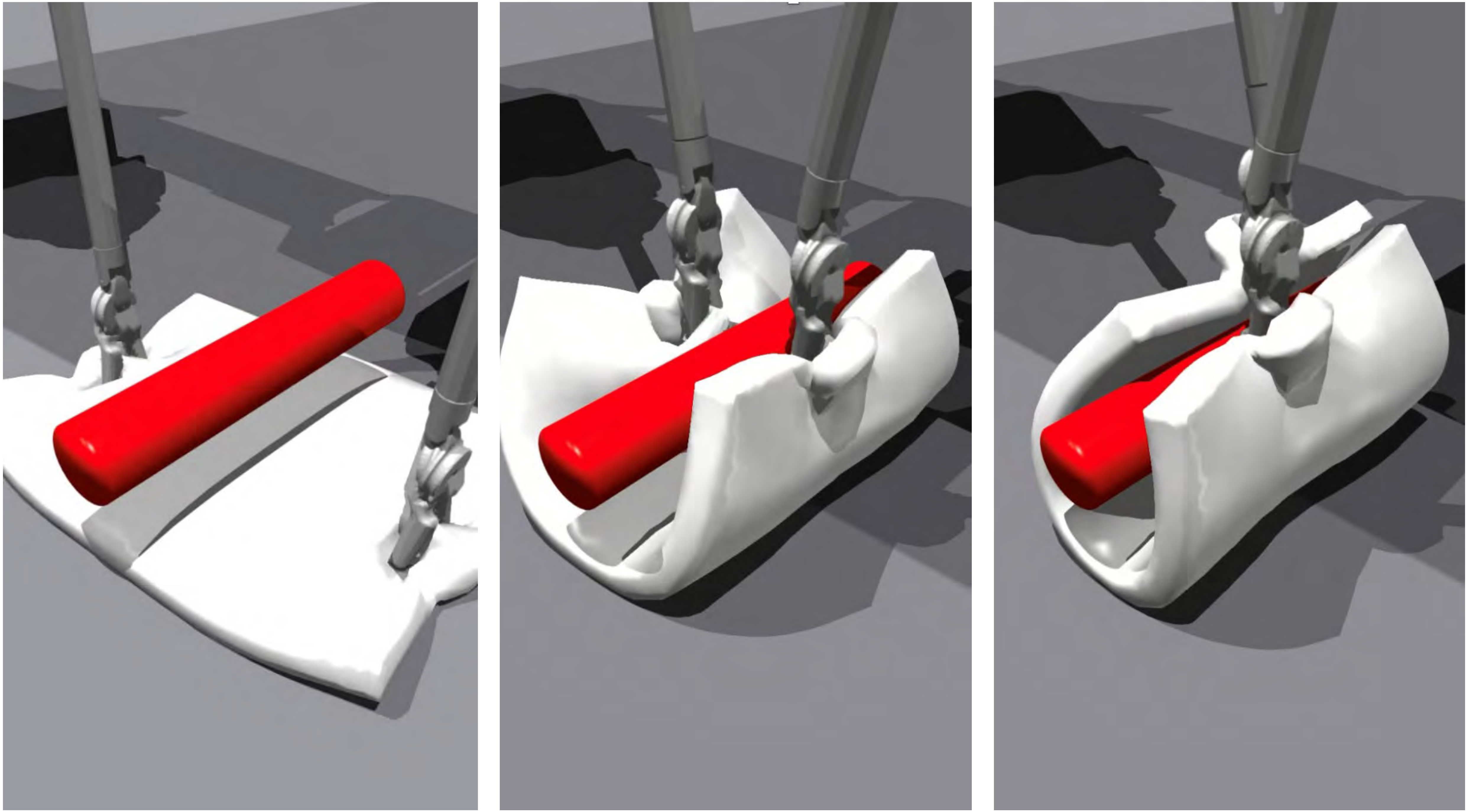










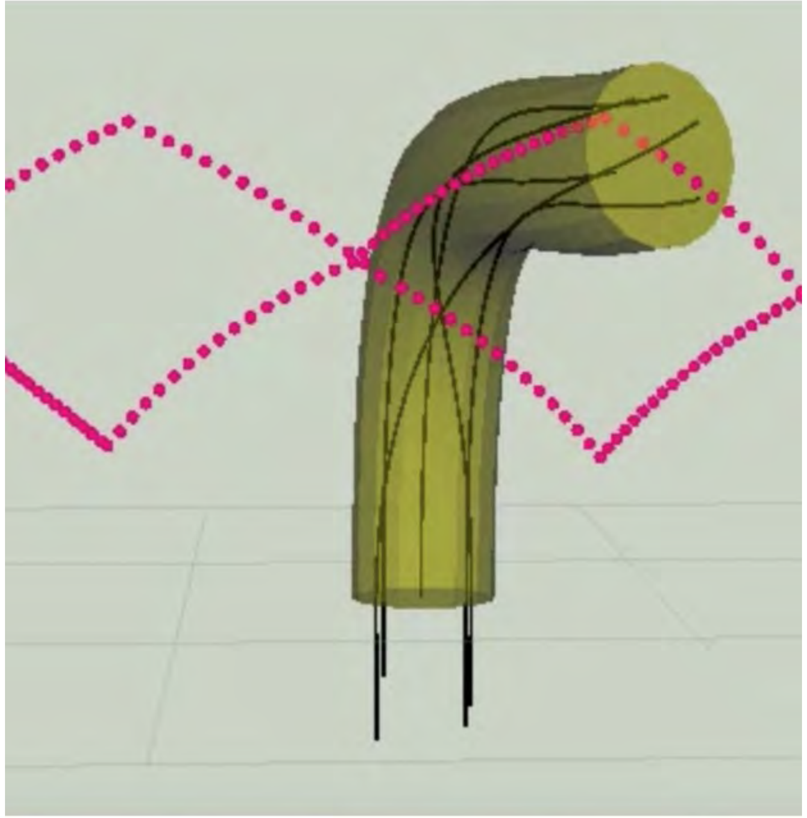
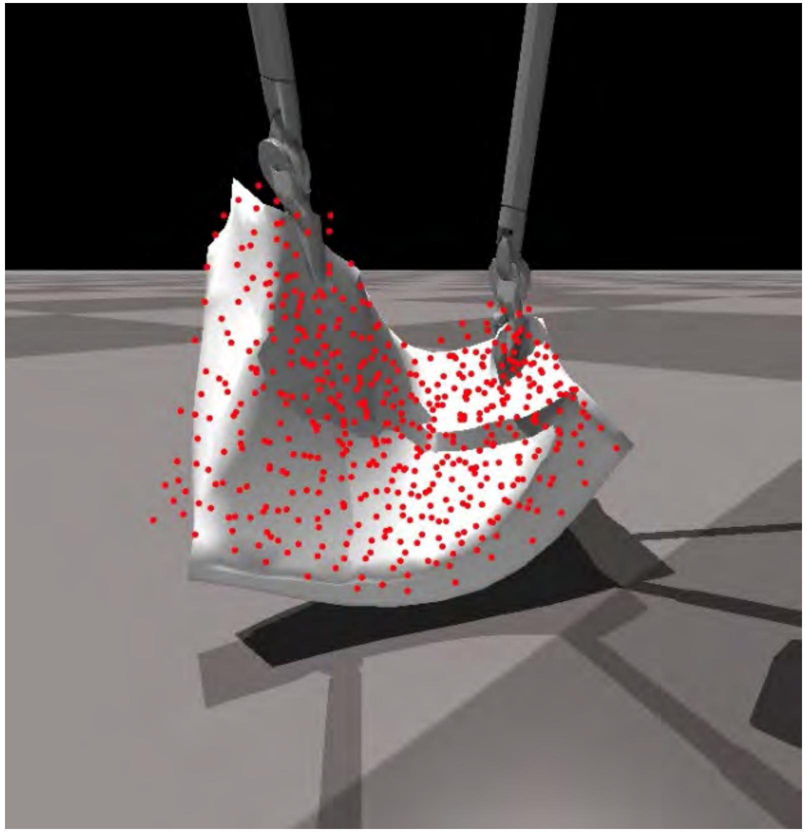
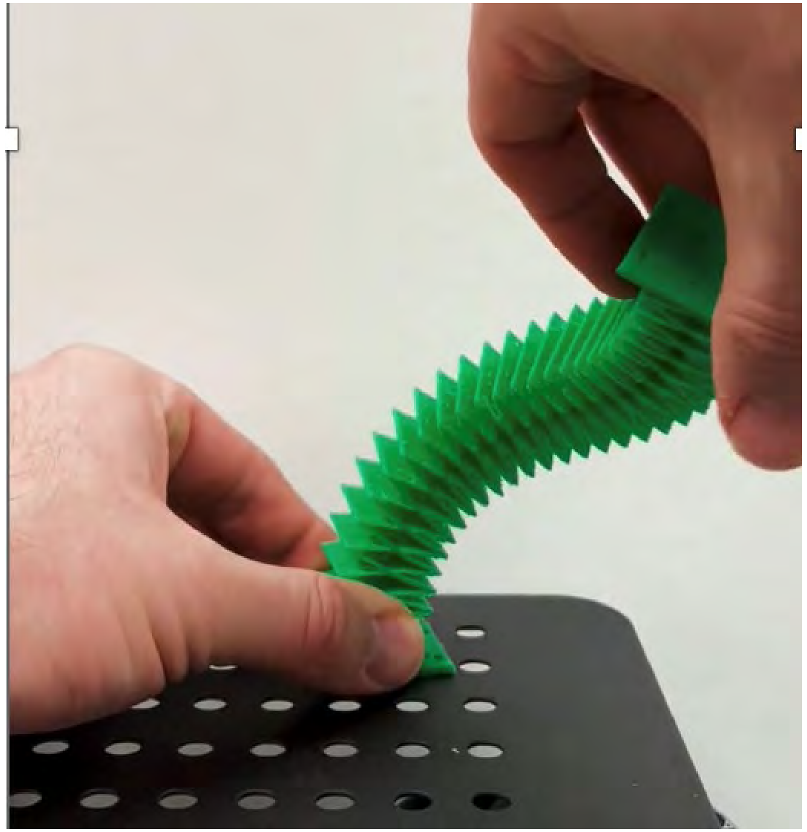


Other University of Utah Robotics/AI/ML

> 15 Core Robotics Faculty

> 20 AI/ML Faculty

Additional strengths in NLP, Computer Vision, Image Processing

Tentacle-Like Continuum Robots and Sub-Task Automation

Alan Kuntz (alan.kuntz@utah.edu)

Assistant Professor, Kahlert School of Computing and Robotics Center

University of Utah

NS&T ML-AI

Big Data   Machine Learning   Artificial Intelligence

Thank you