



Smart-Data for Autonomous Driving and AI

VEHITS 2024

Dr. Arnaud de La Fortelle

heex **Baseline: Intelligence & Data**

Intelligence needs data

- Data as a way to access to reality
- Data opposed to models

Learning from data

- AI as machine learning

Data needs intelligence

- Edge cases and the long tail
- Big Data limitation
- Smart-Data

heex The problem: driving

- Driving looks like a simple problem



heex The problem: driving

- Driving *is* simple
- 2 pedals and 1 steering wheel



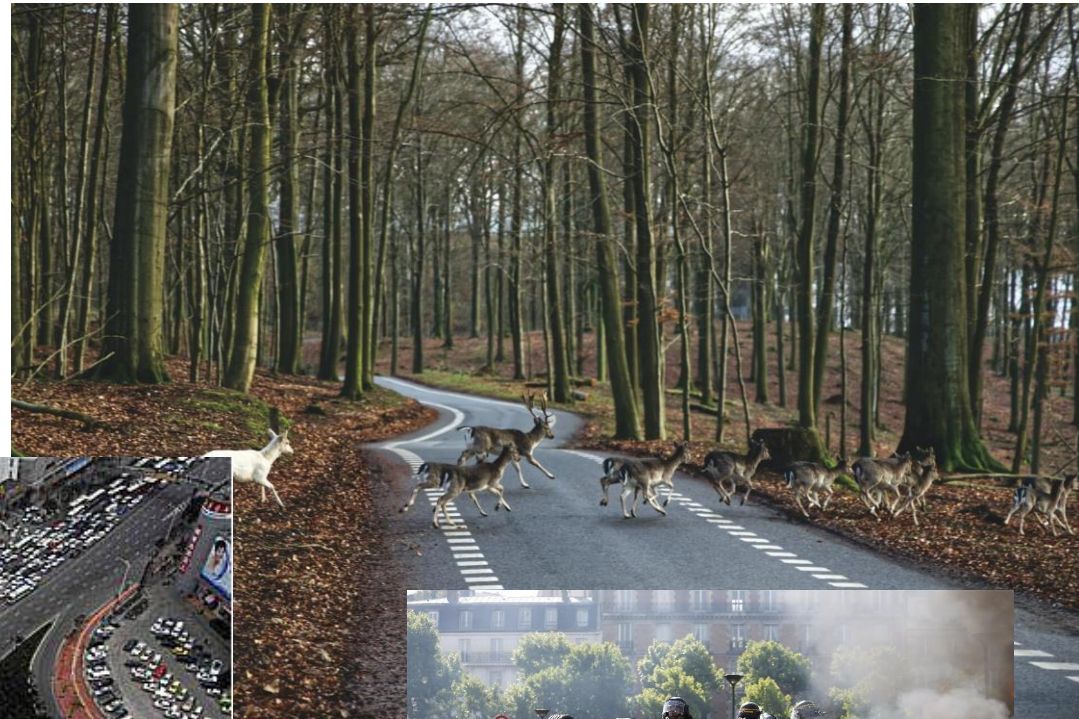
heex The problem: autonomous driving

- Human beings learn to drive in 20h
 - So why is it so difficult to automate?
 - Not enough data?



heex The real world

- The real world is complex and may be wild



heex How to learn driving in a complex world?

- See – Plan – Act paradigm
- Models have failed
 - Very useful within bounds (hypothesis)
 - Not robust to real-world (especially perception)
- AI is taking over
 - Can also make use of models
 - AI is mostly used in perception
 - But is it being used in planning and control

heex Data for AI - Expanding the learning base

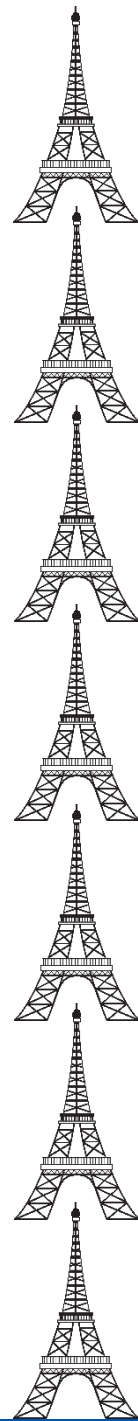
- Deep learning (Lecun)
 - MNIST dataset (1994): 60,000 32x32 images
 - LeNet Deep CNN (1998): 7 layers
- ImageNet: 15 million 300x300 images (~2014)
- GAN (Goodfellow, 2012)
- LLM: ChatGPT (2022)
 - 300 billion words = 500 GB for ChatGPT
 - Who has read 3 million books?
 - Bayesian inference is powerful, Bayes would be happy!
- Now Autonomous Driving?

heex Data from the real world

- 5 TB/h/veh.
 - Mainly because of cameras
- PB/day for a small fleet
- EB/year
 - 1 exabyte = 7 Eiffel towers of hard drives

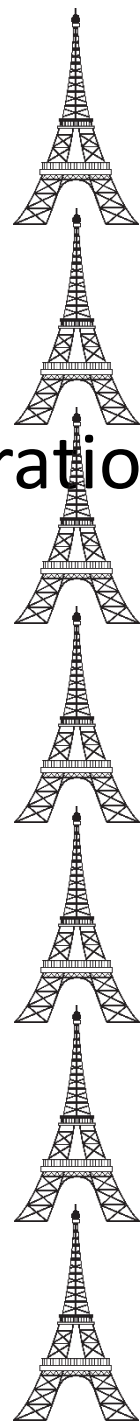
This is BIG DATA

- Not sustainable & not useful



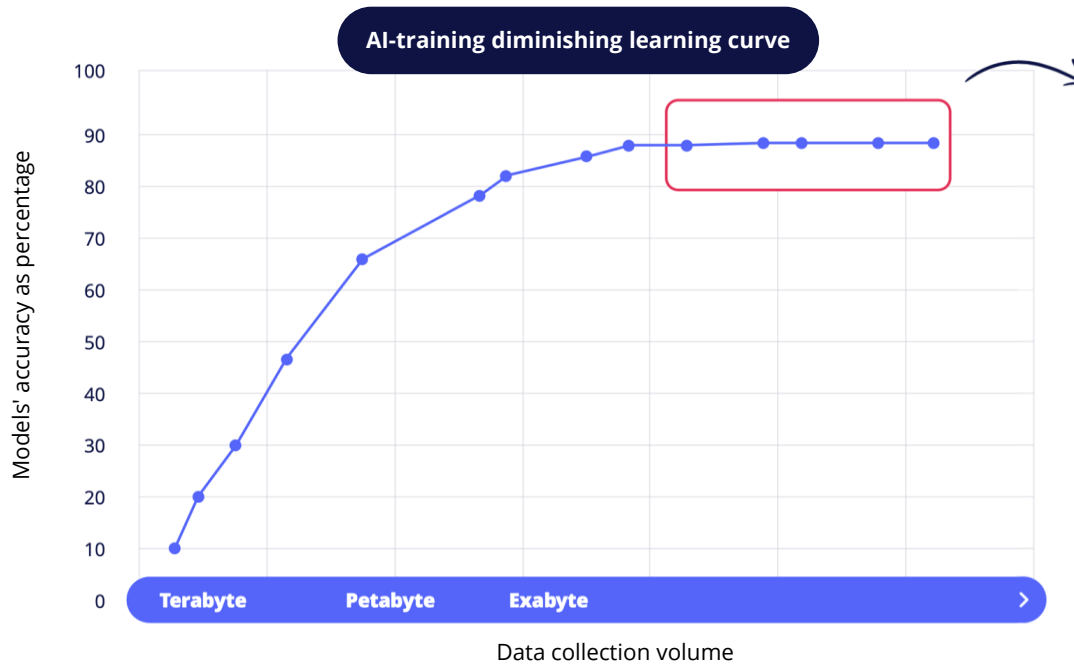
heex Impact of Big Data

- Heex is committed to decrease the impact
- Data is accessed mostly within 72h of generation
- Smart-Data can:
 - Remove unnecessary data: 99%
 - Moving data to cold storage: 92% reduction
 - 1 exabyte = 7 Eiffel towers of hard drives
- Estimated impact: ~1 000 ton CO₂/year



heex Do we need all this data?

The limitations of Big Data to deploy and scale promising AI applications



R&D activities start to reach a plateau as **most of the data collected becomes useless** to learn from new situations (the “edge cases”) and further improve models.



Ever-increasing collection of useless data



Costly, time-consuming, and unsustainable workflow



Delayed time-to-market and failing product releases

heex Smart-Data

- Smart-Data is the synthesized most pertinent data
 - As opposed to Big Data
- What is the relevant data?
 - What do we need and want to learn?
 - Learning everything is not possible, nor useful
 - The Big Data strategy is broken.
- The answer to that is Smart-Data
 - data that is relevant, interoperable, light, real-time, easy to process, and collaborative.
- Build datasets using events

heex Heex solution



SDK

Deployed at the edge or on the cloud to filter and extract Smart Data with a trigger algorithm

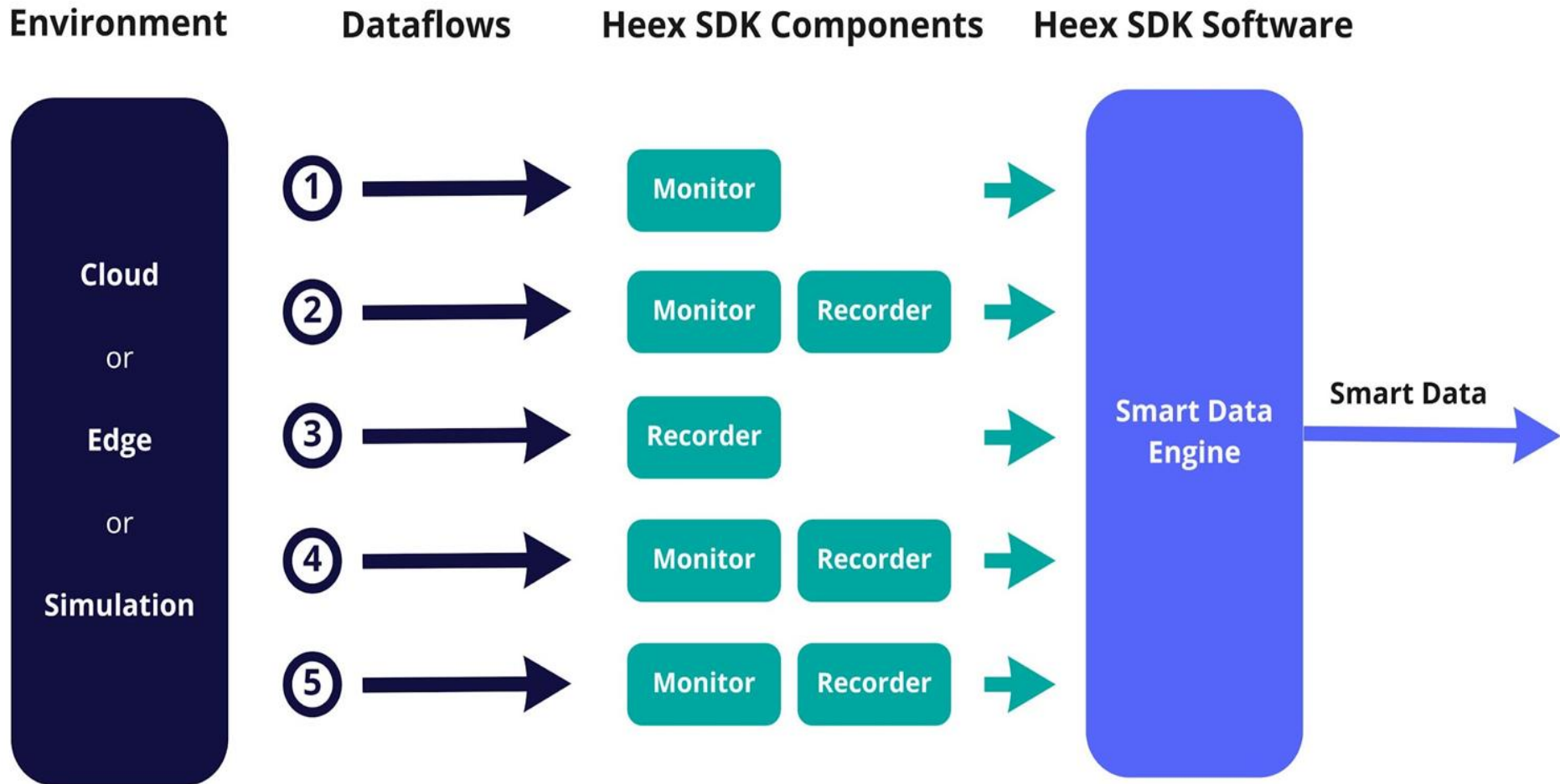
API

Dedicated micro-services backend per client, automating Smart Data triage and distribution

WEB PLATFORM

Web services to access, visualize, manage, and collaborate on Smart Data

heex Events generation



heex **Practical implementation with Carla**

- Digital twins are common
 - Heex interfaces easily with such systems
- Why Carla?
 - A huge success in the community
 - Supports ROS (Carla-ROS bridge)
- Easier to produce massive datasets
 - Can reach TB for sessions
 - Events are ways lighter (2 GB for 20s)

heex 1 sequence in Carla

- Guess what is the Trigger (i.e. the condition)



heex Detailed information to analyze

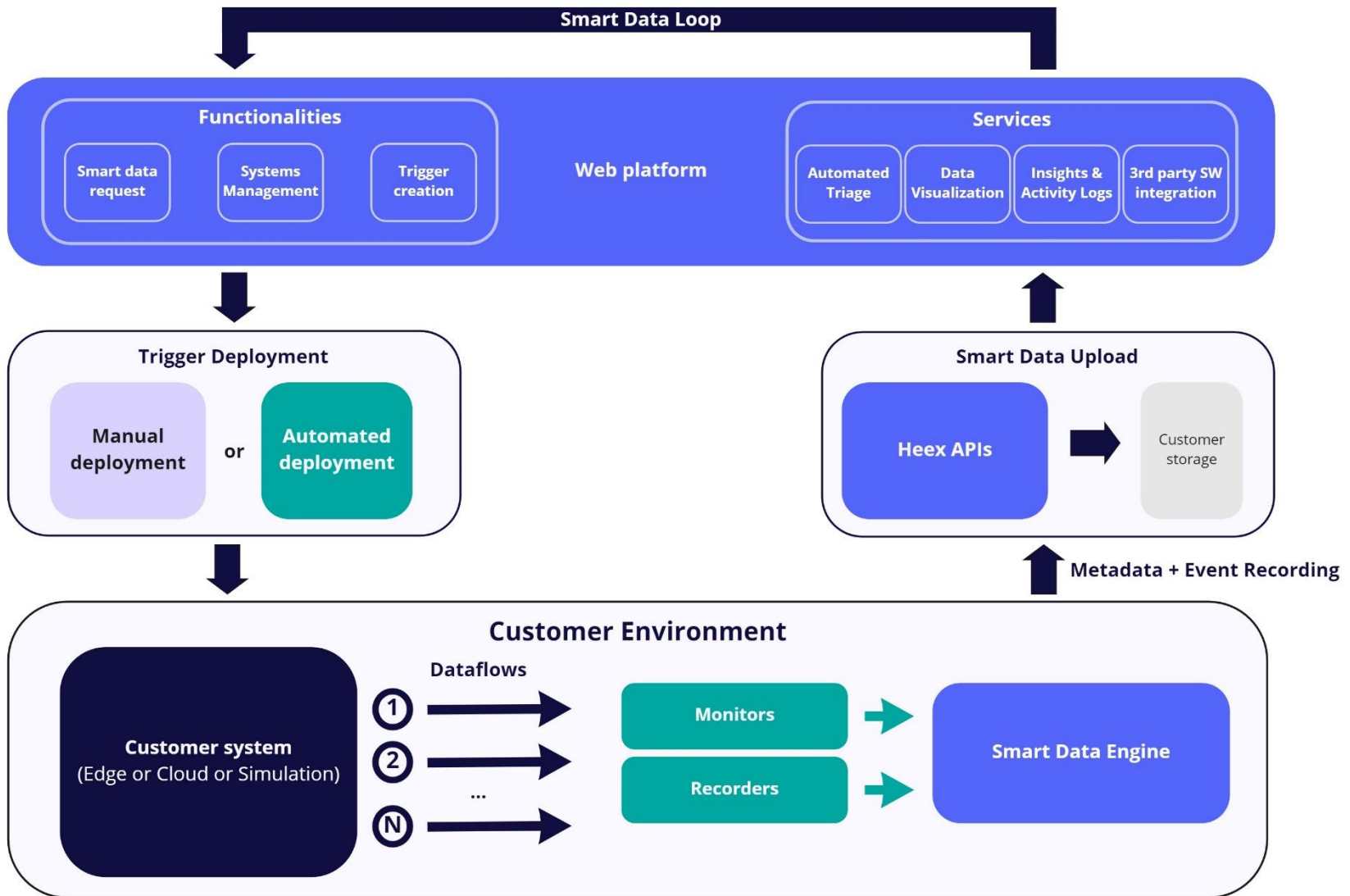
- More than videos
 - Metadata
 - Other contextual data (recorded signals)

```
rosvbag2_bagfile_information:  
  version: 4  
  storage_identifier: sqlite3  
  relative_file_paths:  
    - rosvbag2_2024_03_26-17_16_39_0.db3  
  duration:  
    nanoseconds: 137084801748  
  starting_time:  
    nanoseconds_since_epoch: 1711469799396995460  
  message_count: 10483  
  topics_with_message_count:  
    - topic_metadata:  
      name: /carla/control  
      type: carla_msgs/msg/CarlaControl  
      serialization_format: cdr  
      offered_qos_profiles: ""  
      message_count: 0  
    - topic_metadata:  
      name: /carla/ego_vehicle/vehicle_info  
      type: carla_msgs/msg/CarlaEgoVehicleInfo
```

heex Smart-Data Strategy

- Plan your data collection
 - What is relevant? → Triggers
 - Where to get the information → Systems
- Implement & deploy agents in systems
- Run & gather Smart-Data
 - Automatically generated datasets (1 per Trigger)
- Analyze
- Refine

heex Smart-Data loop



heex Data sharing, trust & cooperation

- Triggers build datasets
 - There can be many of them (hundreds & more)
 - Some can be shared, some not
 - Data is key to evaluation (positive or negative)
- Build trust
 - Transparency is key
- Heex works also with transportation authorities
 - ADS project (US)
 - High expectation of cooperative & automated mobility

heex Conclusion

- Even complex systems require control
 - Managing traffic, behaviors...
- Models become sophisticated
 - E.g. bicycle behaviors
- Huge amounts of data is necessary
 - But all the data is not useful
 - Or not useful to everybody
- Heex Technologies brings its contribution
 - Smart-Data platform

heex 

Thank you!