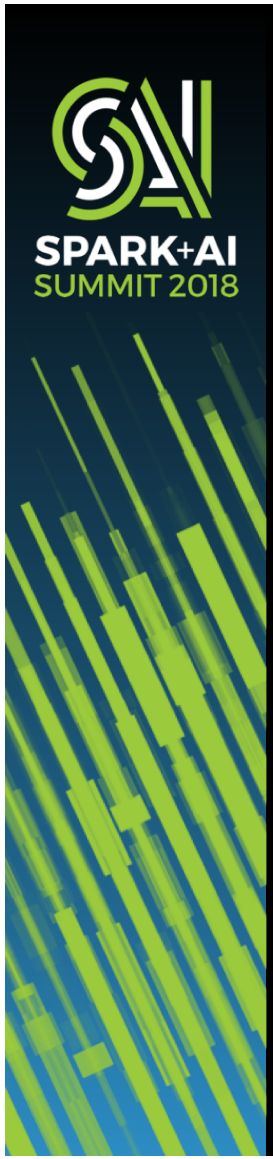



# A Machine Learning Approach to Time-Sensitive Data Analysis

**Anthony Kim**
Staff Data Scientist
Samsung Research America

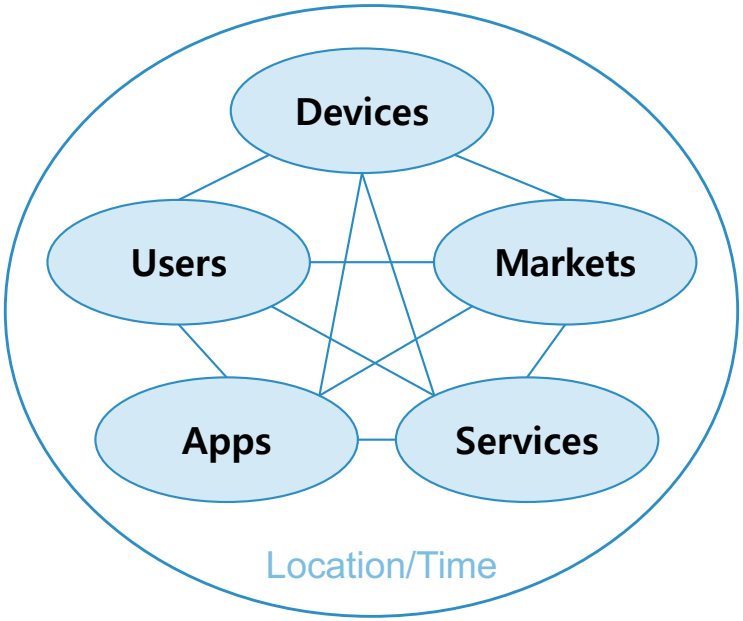
**#Ent4SAIS**

# Acknowledgement

| Machine Learning | Feature Engineering | Management |
|---|---|---|
| Jaemun Sim | Byeonjin Kim Chanmuk Kim | Kwanghyun Koh |
| Sai Kiriti | Jinbeom Lee Wonkeun Oh | VP Chul Lee |

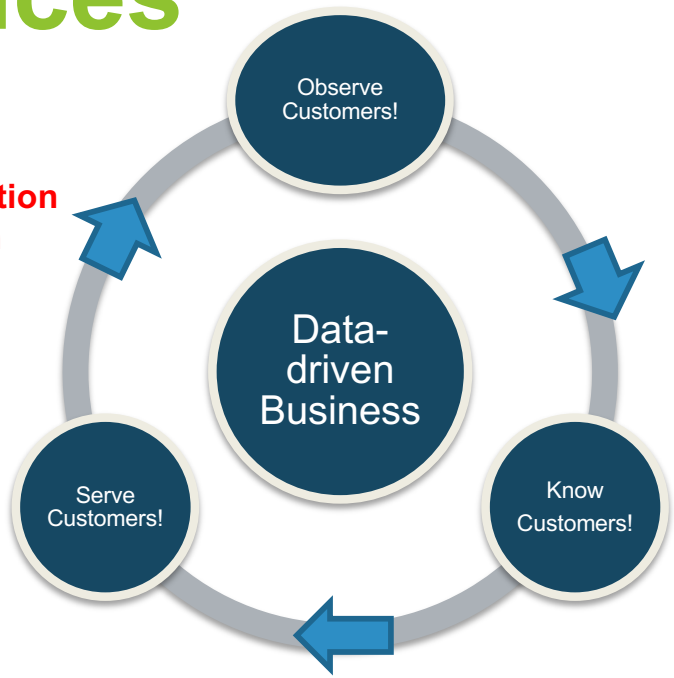
# Big Picture

**Turn Big data into Smart Data via Audience Science to Drive Value Creation**




- Provide **executives** with **actionable business insights**
- Provide **engineering teams** with models for customers' **better product experience**
- Provide **corporate citizens** with **guidance to increase efficiency**

# Smart Devices, Smart Data, and Smart Services




**Machine-oriented Organic Logs:**
- ACR logs
- Smart Hub logs

**Turning MACHINE logs into HUMAN data**

**Human-centric Smart Data:**
- Hierarchical Session Data
- User Profiles / Segments
- Household Profiles / Segments

**Building HUMAN behavioral models**

**Outcome from Machine Learning & Data Mining Models:**
- Targeted Ads / Marketing:
- Personalized Recommendation on:
  - Connected Life
    - Home, Auto, Mobile, …
  - Smart TV Services
  - Contents
  - Consumer Goods

**Feeding HUMAN reaction back into observation**

# Sessionization of Viewing History

- Combine all repetitive log datapoints as one time block if the state remained the same

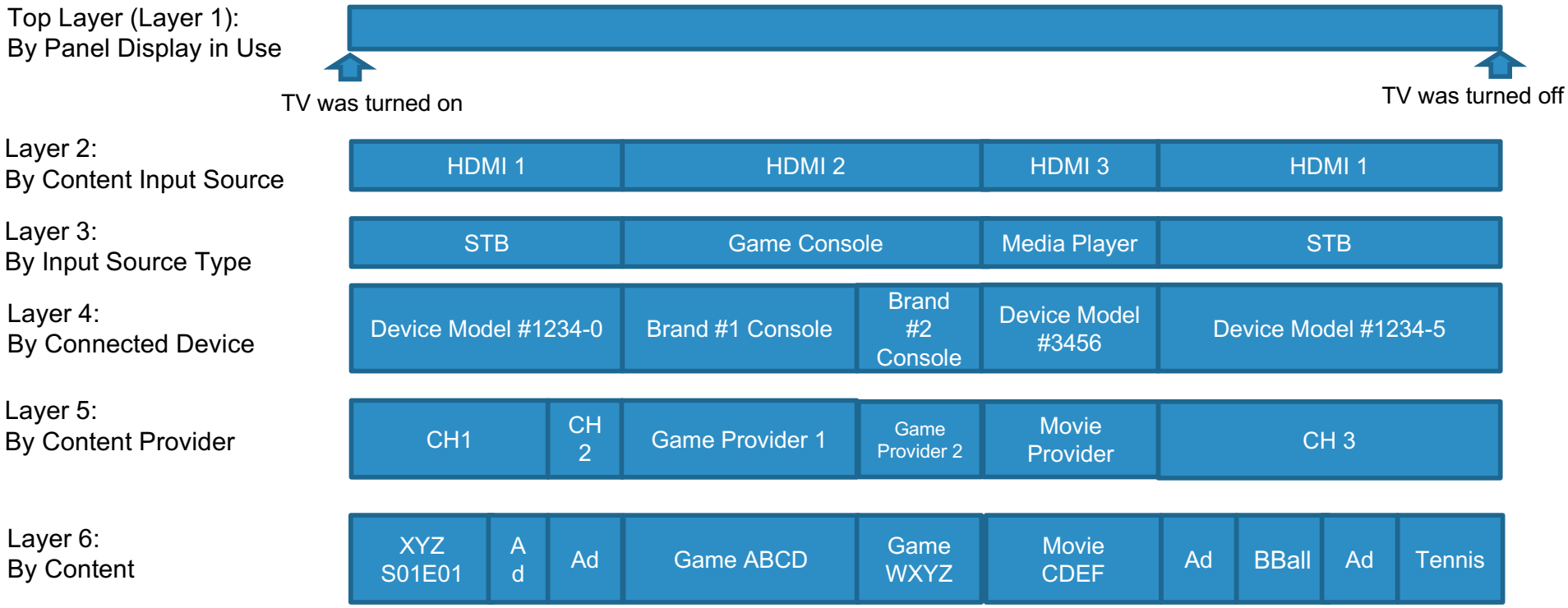
8:00pm Anthony Kim watched “XYZ-TV-SHOW Season 1 Episode 1” on Channel 1
8:01pm Anthony Kim watched “XYZ-TV-SHOW Season 1 Episode 1” on Channel 1
8:02pm Anthony Kim watched “XYZ-TV-SHOW Season 1 Episode 1” on Channel 1
…
8:30pm Anthony Kim watched “XYZ-TV-SHOW Season 1 Episode 1” on Channel 1

8:00pm – 8:30pm Anthony Kim watched “XYZ-TV-SHOW Season 1 Episode 1” on Channel 1

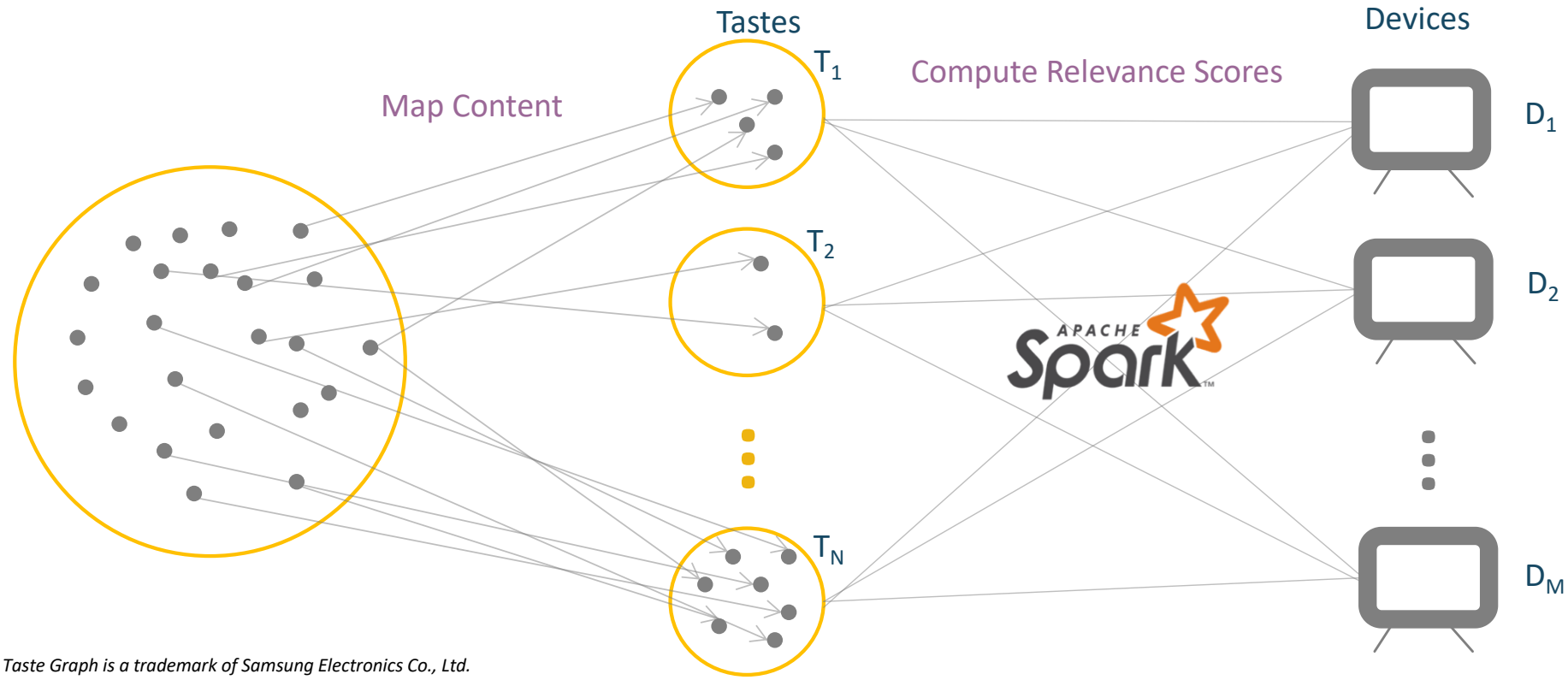
This is one program watch session

*Apache Spark, Spark and the Spark logo are trademarks of The Apache Software Foundation*

# Hierarchical Sessions



# Samsung Taste Graph ®

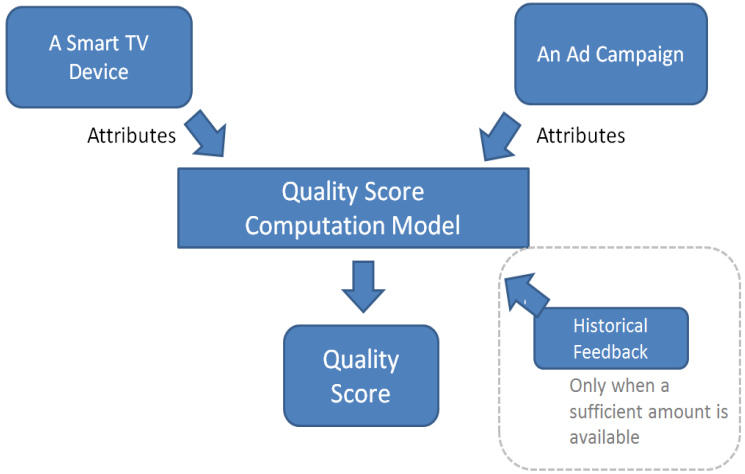



*Samsung Taste Graph is a trademark of Samsung Electronics Co., Ltd.*

# Time Units

- Daily Time Slices (48 Time Slices/Day)
- Day of Week (7)
- Weekdays (5) and Weekend (2)
- Dayparts (8)
  - Prime Time, Late News, Late Fringe, Post Late Fringe
  - Morning, Daytime, Early Fringe, Prime Access

❖ Proprietary *Time-Sensitive* Taste Graph
❖ Proprietary Quality Score




| | TS01 | TS02 | … | TS47 | TS48 |
|---|---|---|---|---|---|
| Day 1 | | | | | |
| Day 2 | | | | | |
| … | | | | | |
| Day N | | | | | |

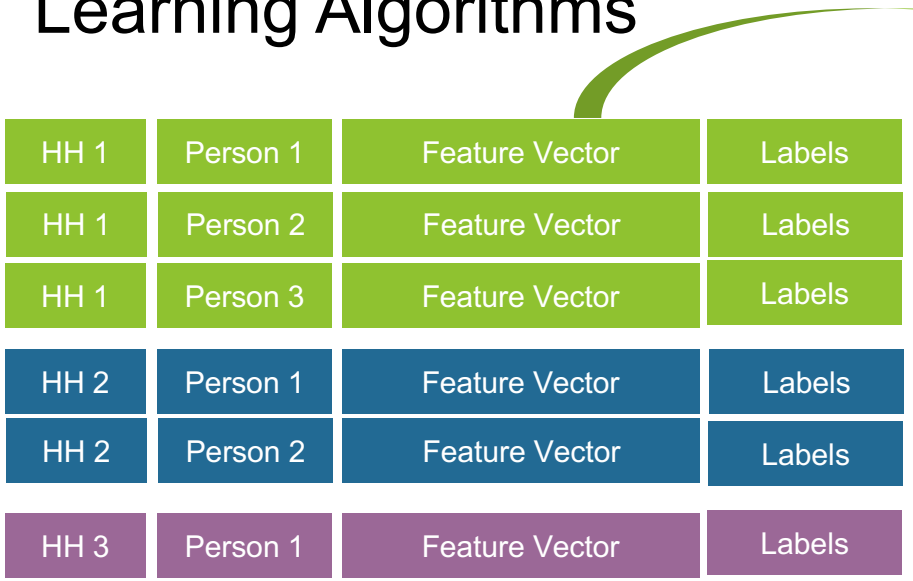
| | Mon | Tue | … | Sat | Sun |
|---|---|---|---|---|---|
| Week 1 | | | | | |
| Week 2 | | | | | |
| … | | | | | |
| Week N | | | | | |

| | Weekdays | Weekend |
|---|---|---|
| Week 1 | | |
| Week 2 | | |
| … | | |
| Week N | | |

| | DP1 | DP2 | … | DP7 | DP8 |
|---|---|---|---|---|---|
| Day 1 | | | | | |
| Day 2 | | | | | |
| … | | | | | |
| Day N | | | | | |

# Household Demographics Prediction

- Combination of General Machine Learning and Deep Learning Algorithms

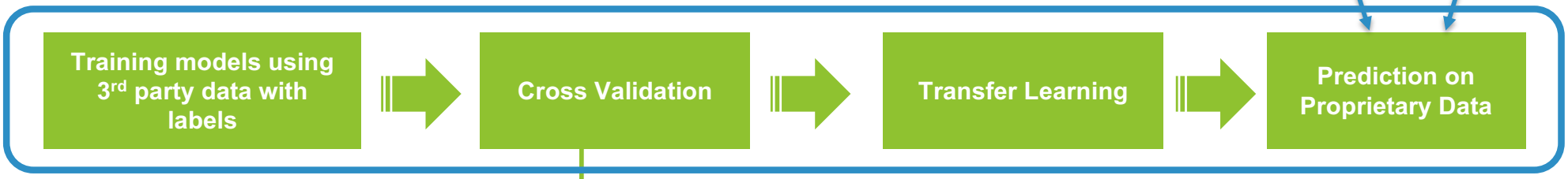



- 7,000+ features
  - Time-sensitive features
  - Keyword features
  - Title features
  - Duration features

# Training and Test Process

on

ACR EPG




| What was predicted | Label | Cross Validated Accuracy |
|---|---|---|
| Gender | Female | 89% |
| | Male | 88% |

| What was predicted | Label | Cross Validated Precision |
|---|---|---|
| # HH Members | 1 | 92% |
| | 2 | 85% |
| | 3+ | 91% |

| What was predicted | Label | Cross Validated Precision |
|---|---|---|
| Age Group | Age 13-17 | 85% |
| | Age 18-24 | 81% |
| | Age 25-34 | 73% |
| | Age 35-44 | 74% |
| | Age 45-54 | 71% |
| | Age 55-64 | 71% |
| | Age 65+ | 80% |

*TensorFlow, the TensorFlow logo and any related marks are trademarks of Google Inc.*

*Horovod is Uber's Open Source Distributed Deep Learning Framework for TensorFlow*

*Databricks, the Databricks logo and any related marks are trademarks of Databricks Inc.*

# Why Databricks?

- Suitable for a wide range of user groups
- Cost-effectiveness
- End-to-end machine learning lifecycle
- Support for distributed deep learning model training

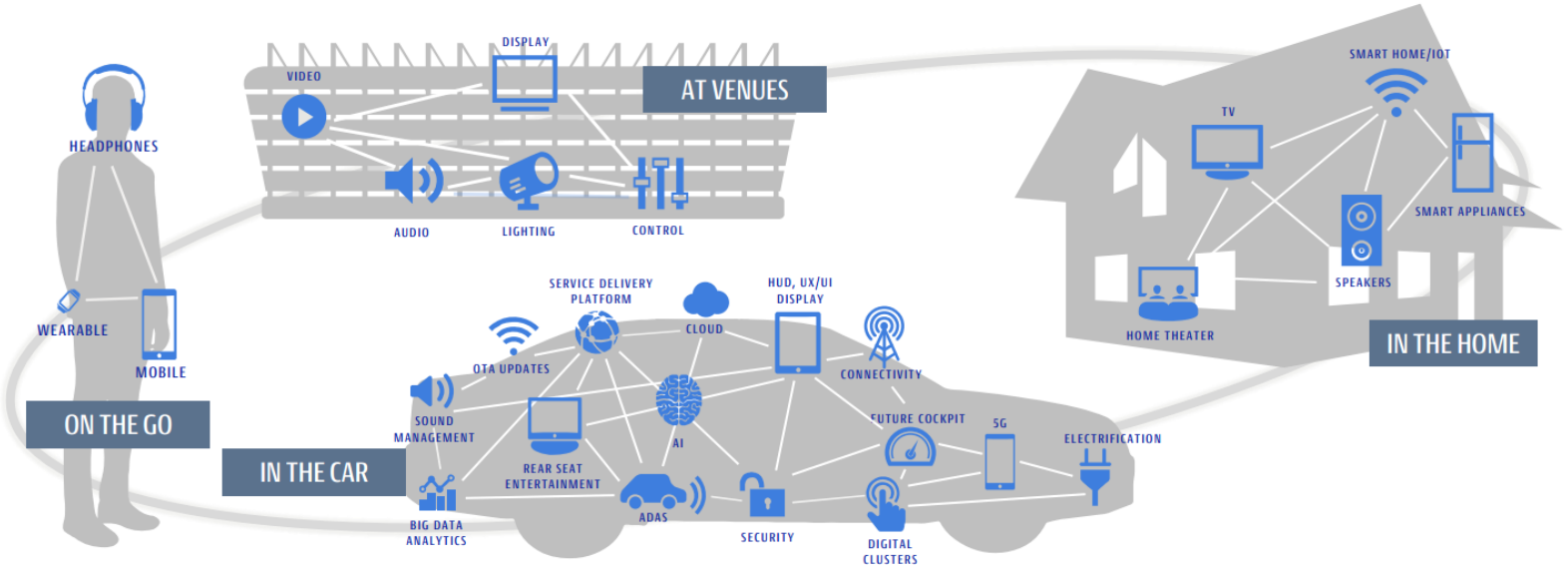
# Summary

- Hierarchical Sessionization
- Samsung Taste Graph ®
- Time-Sensitive Features
  - Proprietary Time-Sensitive Taste Graph / Quality Score
  - Demographics Prediction

Smart Devices, Smart Data, and Smart Services

# Future Work

360-degree View of Whole Processes
(Observation, Profiling, Serving, Feedback, etc.)



# Thank you for your attention!

## Questions?

**hyunwoo.k@samsung.com**

linkedin.com/in/anthony-sra