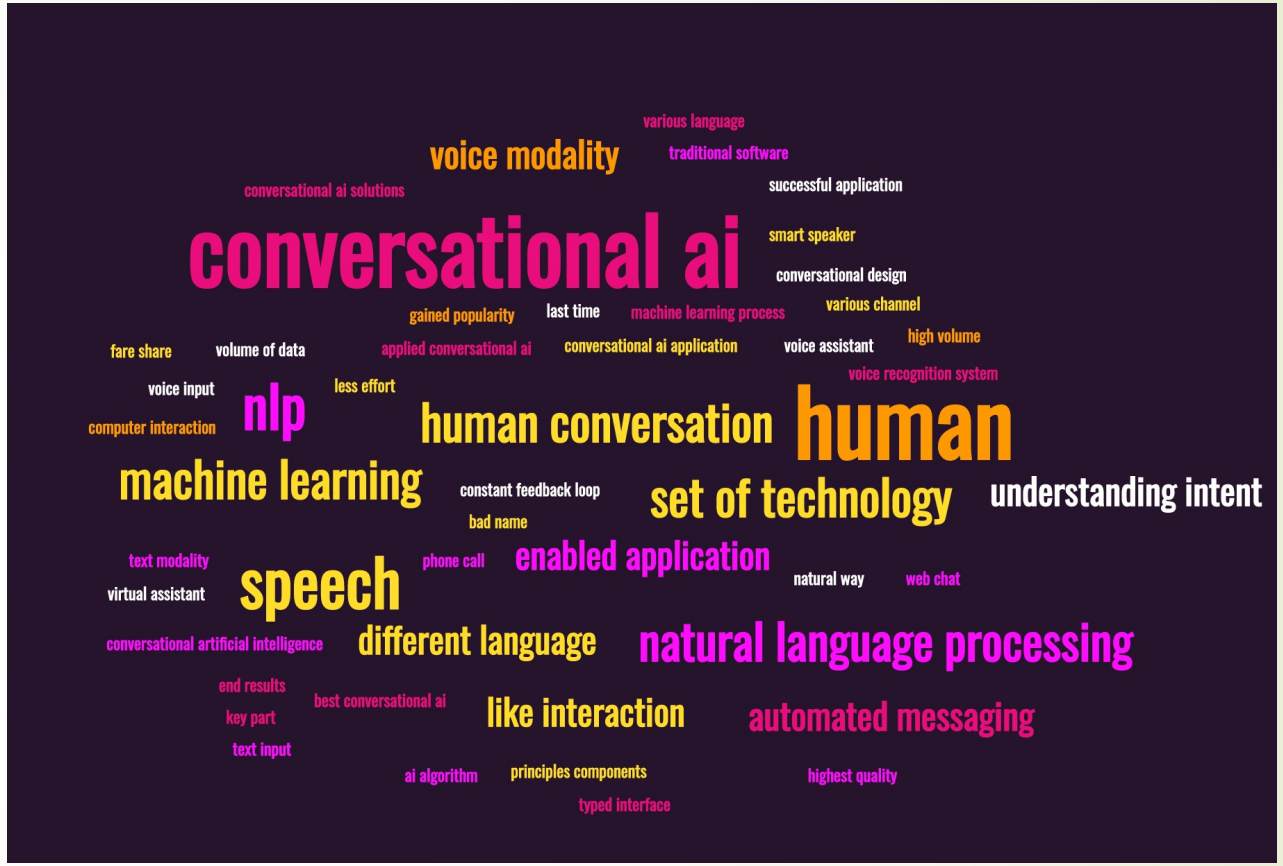
# Conversational AI In Production

## Challenges and Advances

Aravind Ganapathiraju
Uniphore

# What is Conversational AI?

- Deals with Human Conversation
- Set of Technologies
- Speech, Voice
- Natural Language Processing
- Languages
- Intent
- Machine Learning



RIES D FUNDIN
RPRISE AI AND
NS

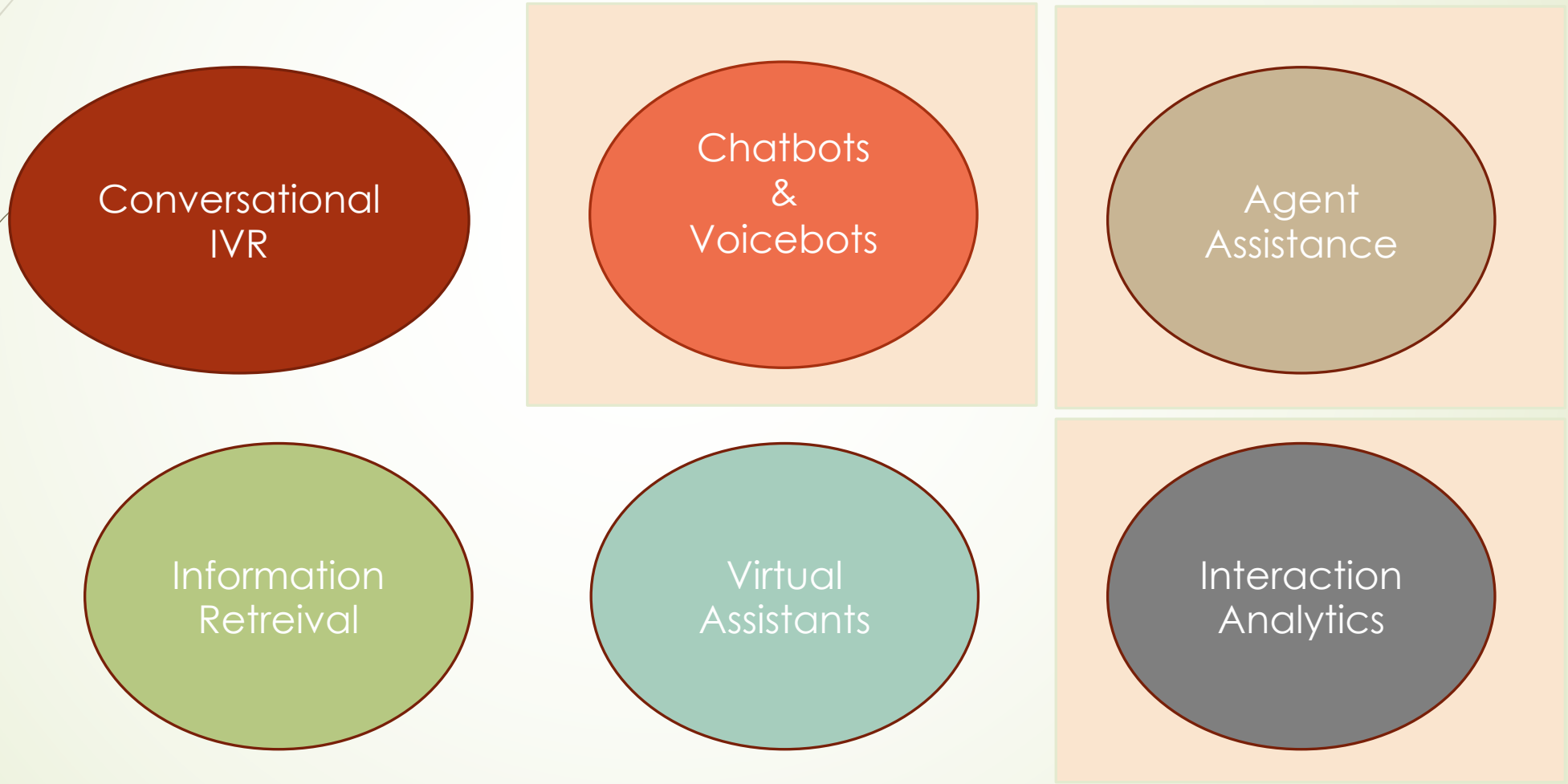
Replicant.ai raises $27M to modernize customer service with

Call center AI firm ASAPP raises $120 mln, valued at $1.6 bln

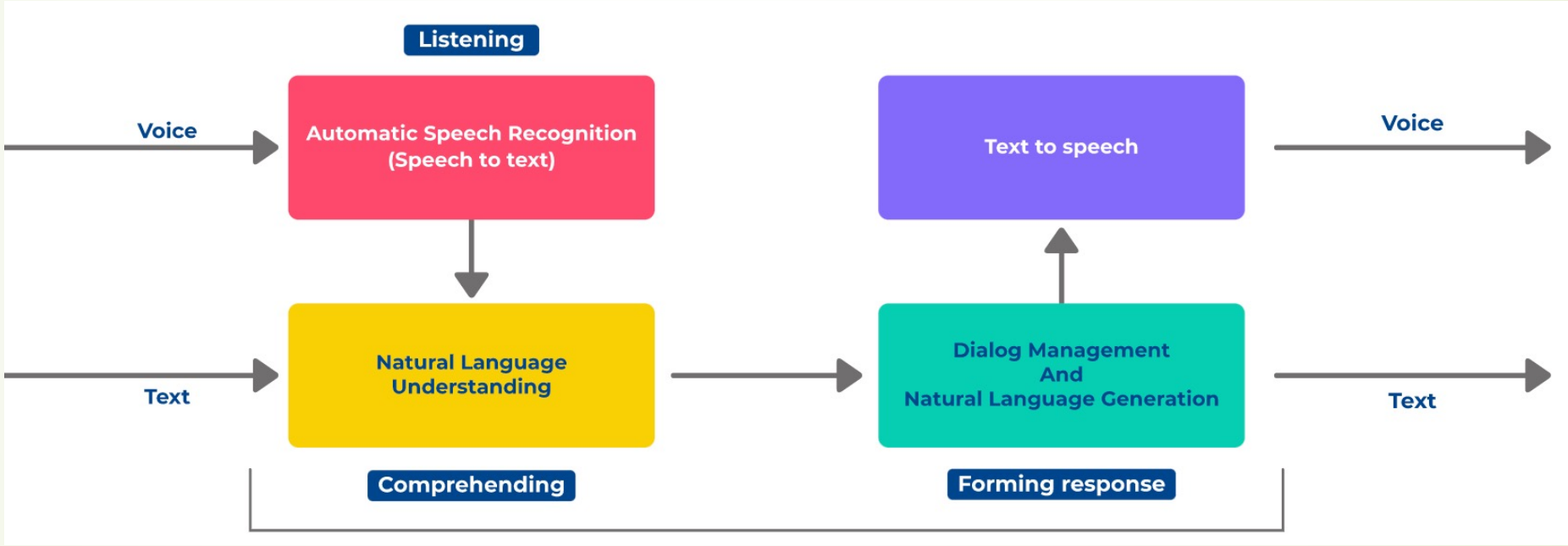
**Observe.AI Raises $54 Million Series B Round; Leverages AI to Transform the Contact Center**

SEPTEMBER 15, 2020 PRESS RELEASE

# Conversational AI – Typical Usecases



# 30000 ft View of Conversational AI



# NLP – Uses and Progress

- Sentiment Analysis
- Natural Language Understanding
- Entity Extraction
- Summarization
- Dependency Parsing
- Dialog State Tracking
- Information Retrieval/Question Answering
- Context and Personalization

- Deep Learning – game changer
- Transformers – Supercharged applications
- Transfer learning makes new applications feasible
- Multilingual opensource models allow growth in new languages
- Evolving opensource community for new languages

# NLP – Uses and Challenges

- **Sentiment Analysis**
- Natural Language Understanding
- Entity Extraction
- Summarization
- Dependency Parsing
- Dialog State Tracking
- Information Retrieval/Question Answering
- Context and Personalization

- Most hyped, limited use?
- Most data available is based on online customer reviews
- Does not translate to conversations – sentiment over long call can vary over time
- Neural approaches only slightly better than rule/keyword based approaches
- Sarcasm, negation, coreferences- all trip the most complex systems

# NLP – Uses and Challenges

- Sentiment Analysis
- **Natural Language Understanding**
- Entity Extraction
- Summarization
- Dependency Parsing
- Dialog State Tracking
- Information Retrieval/Question Answering
- Context and Personalization

- Transformers and HuggingFace have changed the landscape
- Neural approaches allow for jointly optimized models
- Hybrid approaches best suited when intent data is limited
- Active learning pipelines play a vital role in continuous model improvements
- Lack of annotated data, especially, non-English
- Limited focus on conversation structure and sectioning

# NLP – Uses and Challenges

- Sentiment Analysis
- Natural Language Understanding
- **Entity Extraction**
- Summarization
- Dependency Parsing
- Dialog State Tracking
- Information Retrieval/Question Answering
- Context and Personalization

- Spacy and seq2seq models
- Importance of Inverse text Normalization (ITN)
- Language data scarcity
- Robustness to ASR errors
- Entity resolution issues
  - Relative times and dates
  - World knowledge
- Speech dysfluencies
  - Restarts, stutters, repeats
- Multiturn entities

# NLP – Uses and Challenges

- Sentiment Analysis
- Natural Language Understanding
- Entity Extraction
- **Summarization**
- Dependency Parsing
- Dialog State Tracking
- Information Retrieval/Question Answering
- Context and Personalization

- Extractive summarization most common
- Abstractive summarization showing promise
- Template based summaries most accurate and functional
- Excessive reliance on entity resolution and annotations
- Limited use of audio features for summaries

# NLP – Uses and Challenges

- Sentiment Analysis
- Natural Language Understanding
- Entity Extraction
- Summarization
- Dialog State Tracking
- Dependency Parsing
- Information Retrieval/Question Answering
- **Context and Personalization**

- External context rarely infused into NLP pipeline
  - Prior history
  - Information from other modalities
- Usecase: If an email was sent about top-up loans, can we prime the ASR for inbound call for loans?
- How do you scale user-specific models?
- Featurization of diverse contexts
  - Short and long term contexts
  - Local context for entities

# ASR – SOTA and Advances

- Kaldi and other opensource tools
  - Tons of deployments
  - Well defined recipes
  - Known compute needs
- Next generation E2E Neural Systems
  - ESPNet
  - Nvidia Jasper
  - Mozilla Deepspeech
  - Facebook wav2vec
- Data availability
  - Mozilla Commonvoice
  - Librispeech
  - Highquality CC from Online Video Platforms
  - Indic Language Data
  - Gigaword Corpus
- Sponsored ASR Challenges
  - Igniting low resource language research
  - Data availability
  - Encouraging multilingual systems

# ASR – So What's the Problem?

Audio Quality

Domain Knowledge

Context Unaware

Accents

Access to Data for Model Tuning

Unsupervised Learning Still Ineffective

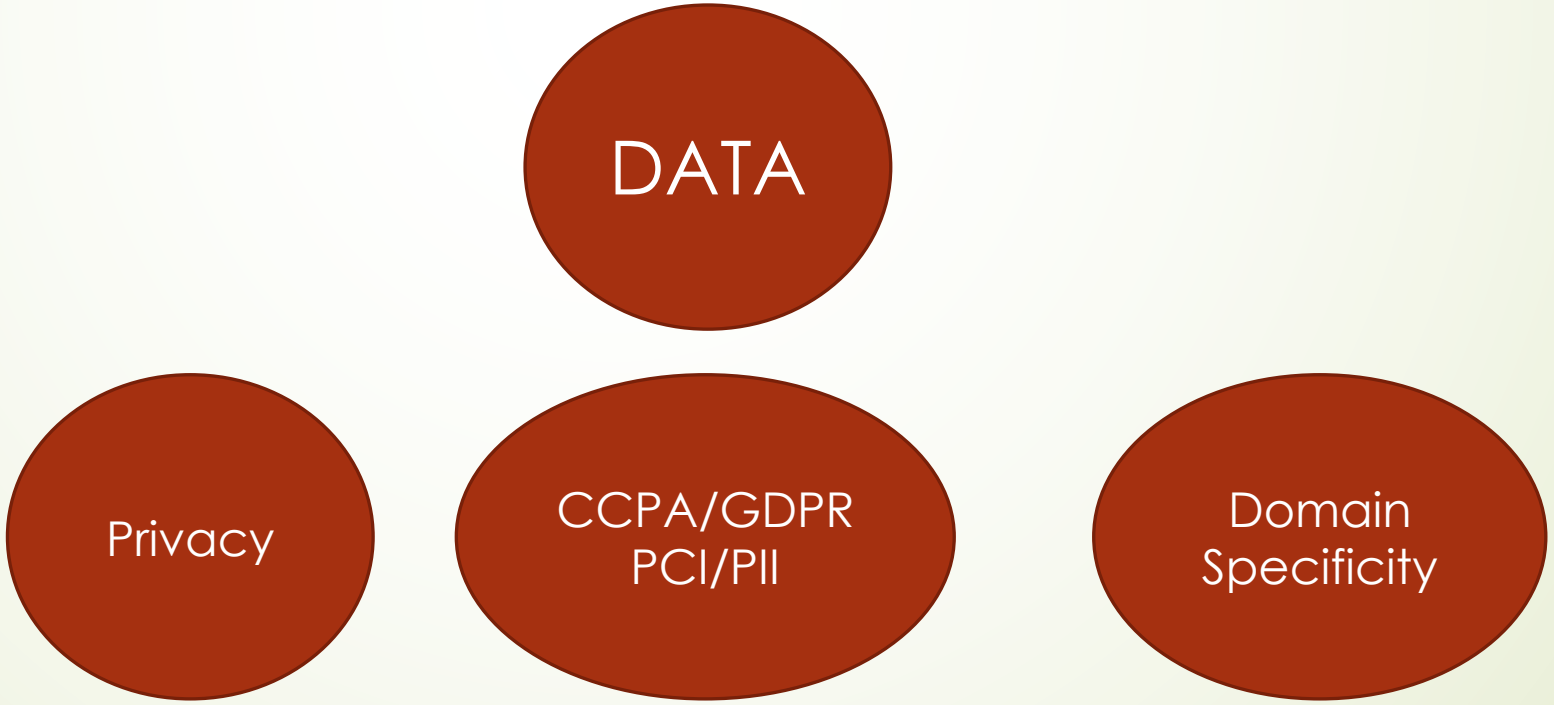
Lack of Evolving Metrics

# ASR – Application Needs

- Compute needs
  - GPUs still expensive
- Latency plays spoilsport for real-time guidance
- Concurrency always a challenge – RTF and memory usage
- Day zero performance still subpar – WER often close to 40% for non-English
- Information sharing between downstream NLP and ASR
  - Context in VAs improves entity recognition
  - Speech Priming helps

# All Set? … Not quite!

- Core technology has advanced
- Compute is available
- Then what's the challenge?
- Primary obstacle is ….



# Data - Reality On the Ground

- Opensource data is not sufficient for enterprise use-cases
- Expensive and time-consuming
- Corpora for languages beyond top-10 are scarce
- Customer data comes with restrictions

Domain Language Models

Customer Specific Models

On-prem *Finetuning*

# More Challenges

- Enterprises ignore effect of data quality and quantity
  - Still use mono instead of stereo
  - Continue to compress data to as low as 12kbps
  - As little as 20hrs of speech for domain knowledge
- Components still optimized individually
  - Need to look at task driven optimization such as ASR tuned for entity extraction
- Dialects, code switching, multilingualism

# Accumulation of Errors

- Language ID wrong – disaster with transcriptions
- Speaker Diarization wrong – disaster with NLP relying on turn identification
- VAD issues – horrible user experience with HCI
- ASR issues – poor transcripts, poor insights
- NLP issues – poor user experience with HCI and unreliable call analytics

**THEY ALL ADD UP**

# Tooling for post deployment - Often ignored

Active Learning

Authoring Help

Feedback Loop

Vocabulary Drift Checks

Model Health Checks

Incremental Tuning with BYOD

# Is Federated Learning a Possibility?

- If we can't get data from customers, lets take training to the customer
- Lots of Interest
- Proven to work with voice assistants
- Challenges with disparate model architectures

**Federated Acoustic Model Optimization for Automatic Speech Recognition**

Conghui Tan[1](✉), Di Jiang[1], Hua
Weiwei Zhao[1], Chaotao Chen[1]
and

**Training Speech Recognition Models with Federated Learning: A Quality/Cost Framework**

Publisher: IEEE Cite This PDF

Dhruv Guliani ; Françoise Beaufays ; Giovanni Motta All Authors

**Federated Learning in ASR: Not as Easy as You Think**

Wentao Yu, Jan Freiwald, Sören Tewes, Fabien Huennemeyer, Dorothea Kolossa

**End-to-End Speech Recognition from Federated Acoustic Models**

Yan Gao, Titouan Parcollet, Salah Zaiem, Javier Fernandez-Marques, Pedro P. B. de Gusmao, Daniel J. Beutel, Nicholas D. Lane

# The Future Is Bright

- Machine learning advancements will continue to break new barriers
  - Leaner transformers
  - Multilingual approaches
  - Finetuning of models with minimal data
- Opensource is more vibrant now, than ever before
- Governmental sponsorship for addressing data scarcity is on the rise
- Cloud adoption by enterprises will allow for CI/CD in AI models
- Inexpensive compute will enable adoption of complex models in production

# Thanks!

Contact me at: aravindganapathiraju@uniphore.com