Platform for Situated Intelligence

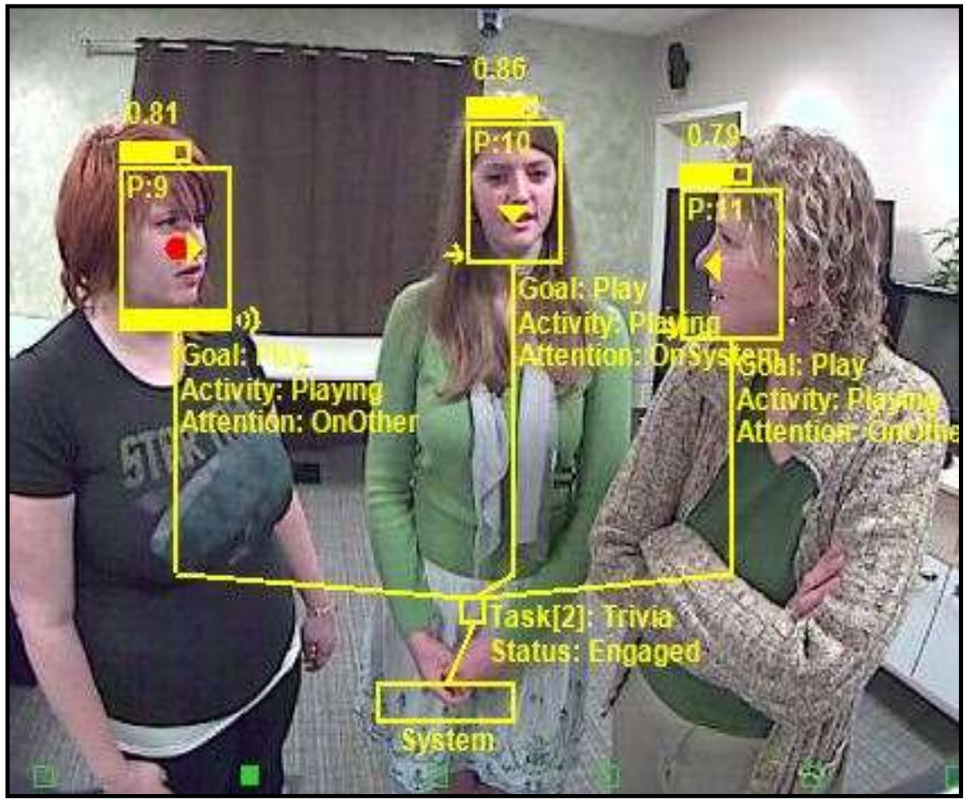
enabling interactive, real-time AI

Sean Andrist, Dan Bohus, John Elliott, Ashley Feniello, Don Gillett, Eric Horvitz, Mihai Jalobeanu, Anne Loomis Thompson, Nick Saw

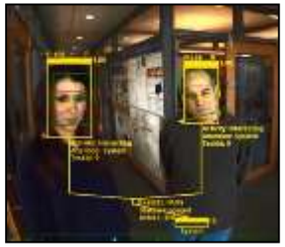
MSR AI, Redmond

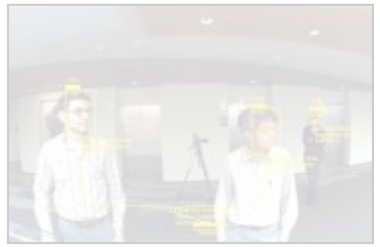
Physically Situated Language Interaction

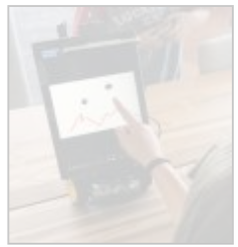
enable computers to perceive and reason about their surroundings and engage in language interaction in physically situated settings



Physically Situated Language Interaction

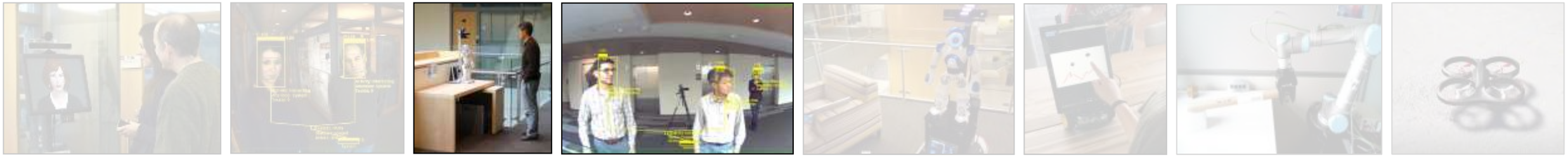






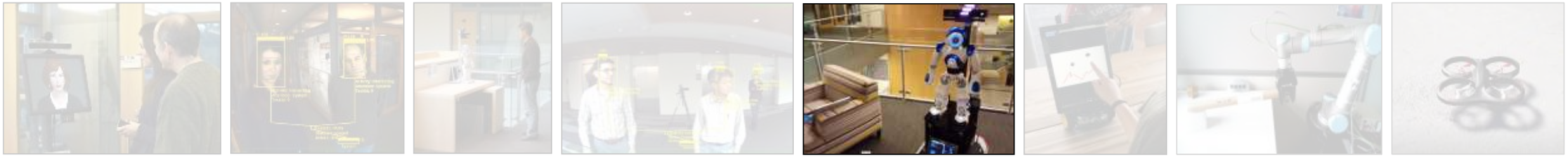
Virtual assistants

Physically Situated Language Interaction



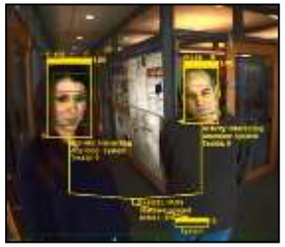
Human robot interaction

Physically Situated Language Interaction



Social mobile robots

Integrative AI

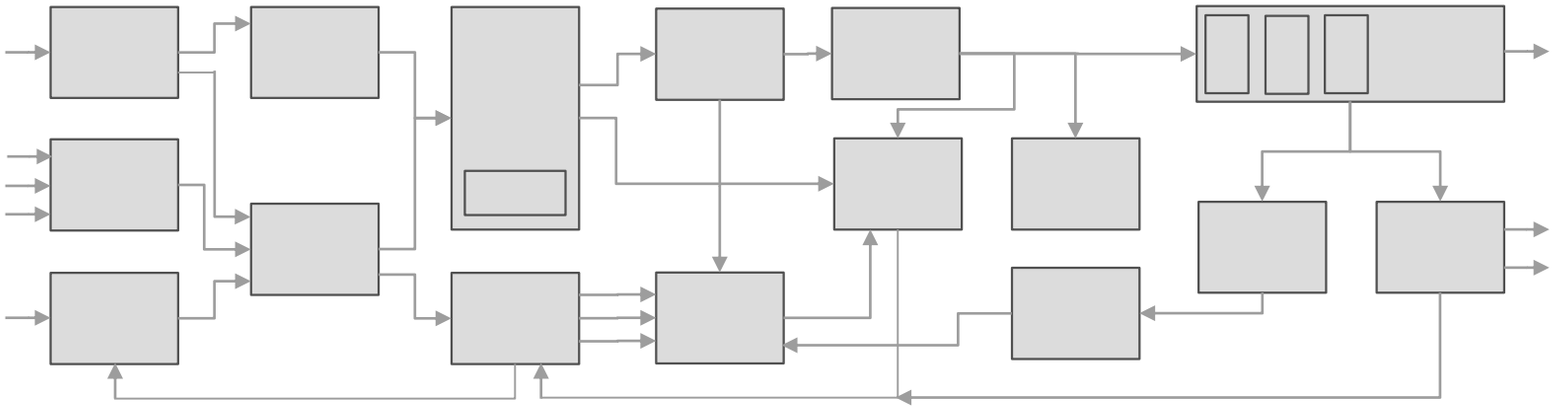


Integrative AI development challenges

Complexity

Missing primitives

Debugging



Microphone array capture
Sound source localization
Speech recognition
Language understanding
Infrared proximity sensors
Badge sensors
Face detection and tracking
Head-pose tracking

Facial feature tracking
Face identity recognition
Gender detection
Attention models
Engagement models
Turn-taking models
Behavioral control

Dialog management
Natural language generation
Speech synthesis
Avatar synthesis
Robot motion control
Floor-plan models
User models

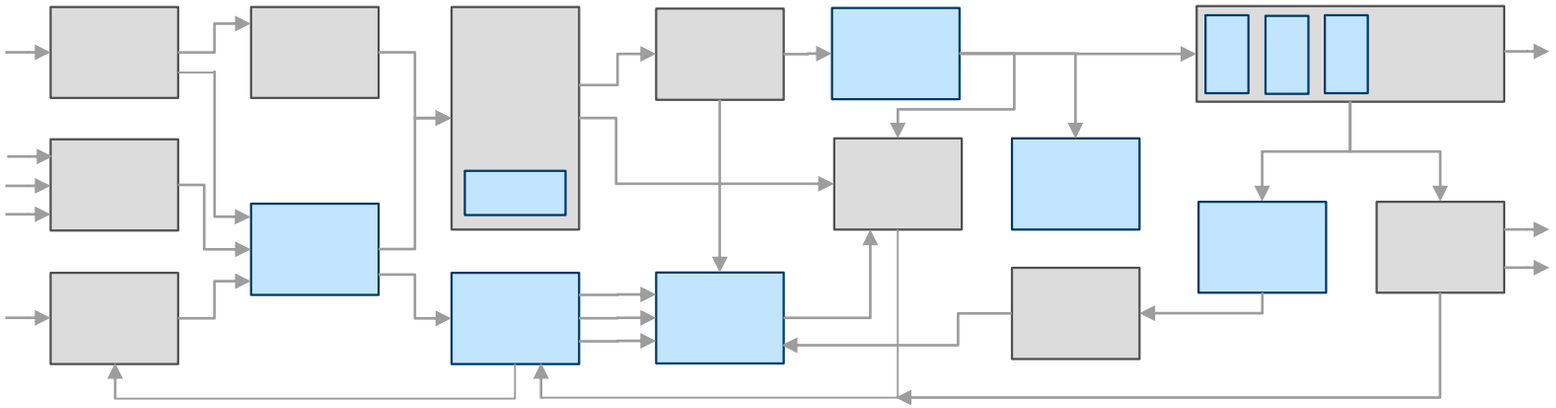
Integrative AI development challenges

Complexity

Missing primitives

Debugging

Machine learning



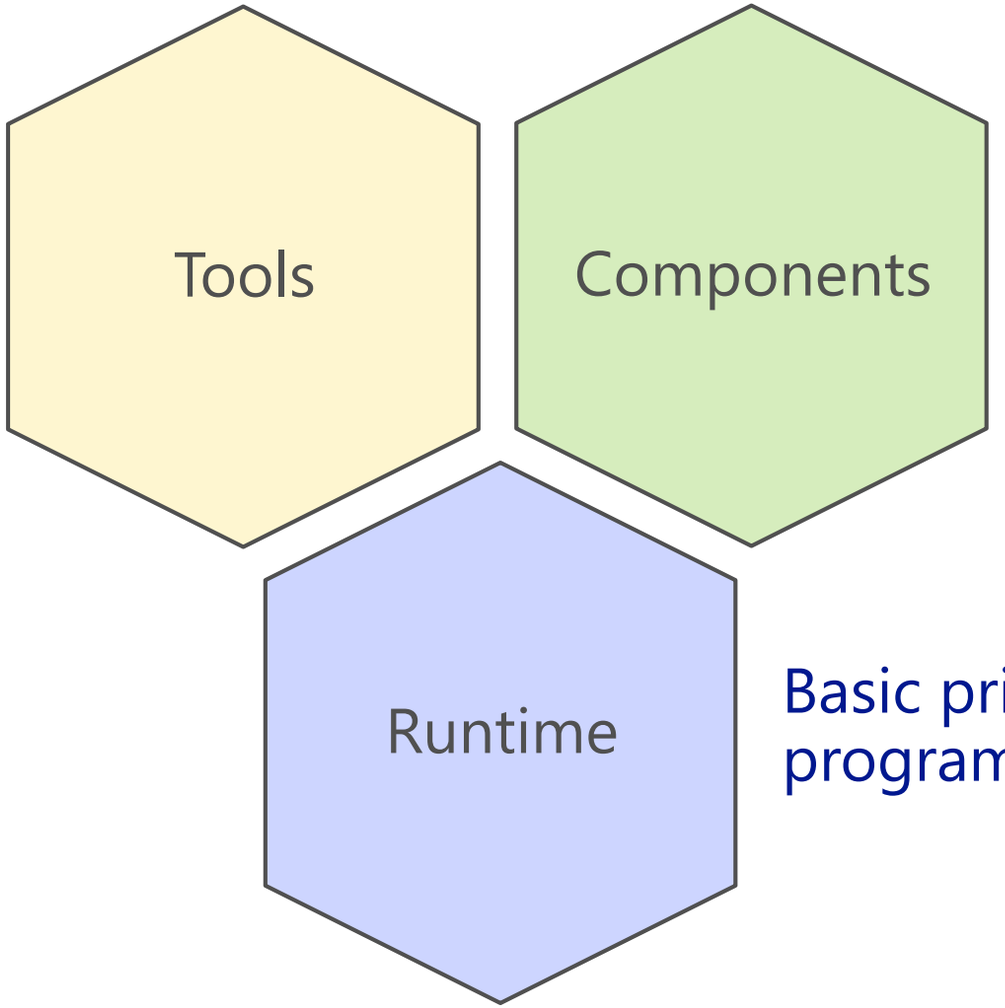
Microphone array capture
Sound source localization
Speech recognition
Language understanding
Infrared proximity sensors
Badge sensors
Face detection and tracking
Head-pose tracking

Facial feature tracking
Face identity recognition
Gender detection
Attention models
Engagement models
Turn-taking models
Behavioral control

Dialog management
Natural language generation
Speech synthesis
Avatar synthesis
Robot motion control
Floor-plan models
User models

Platform for Situated Intelligence

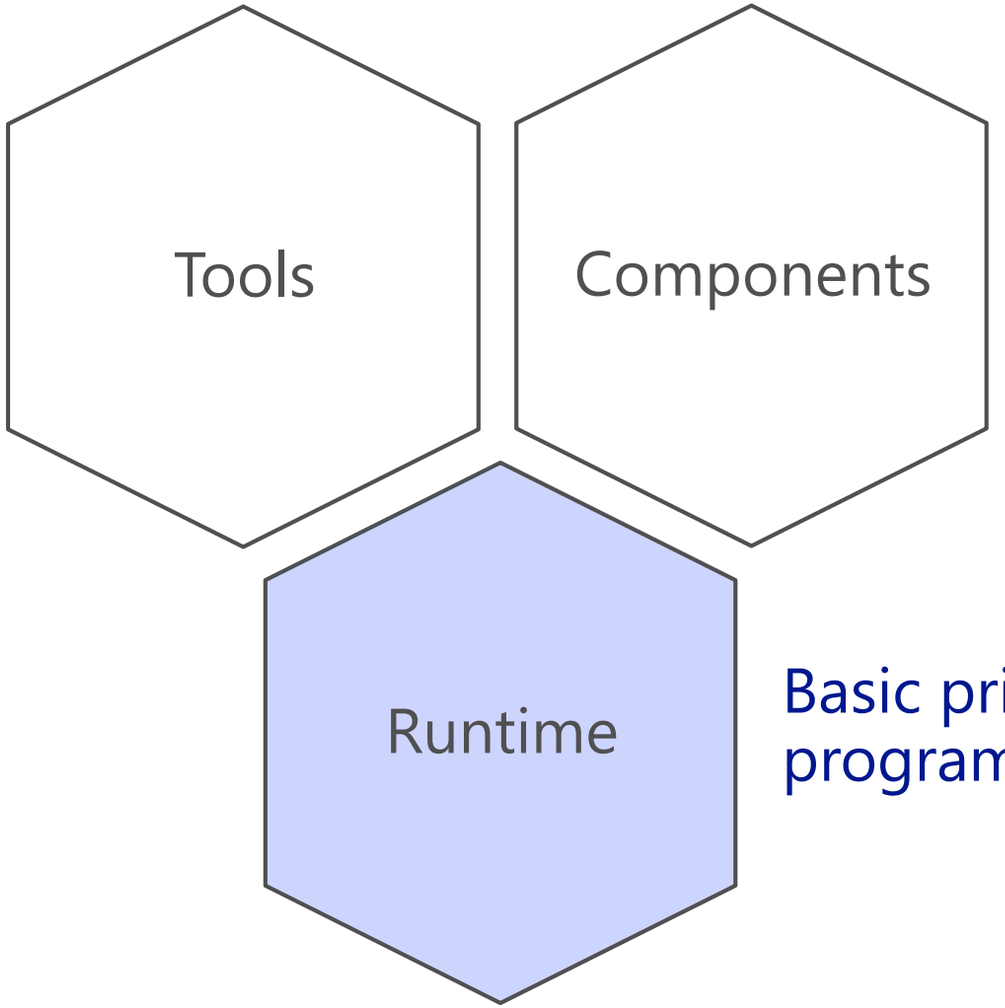
Debugging
Visualization
Machine Learning



Repository of
AI technologies

Basic primitives and
programming model

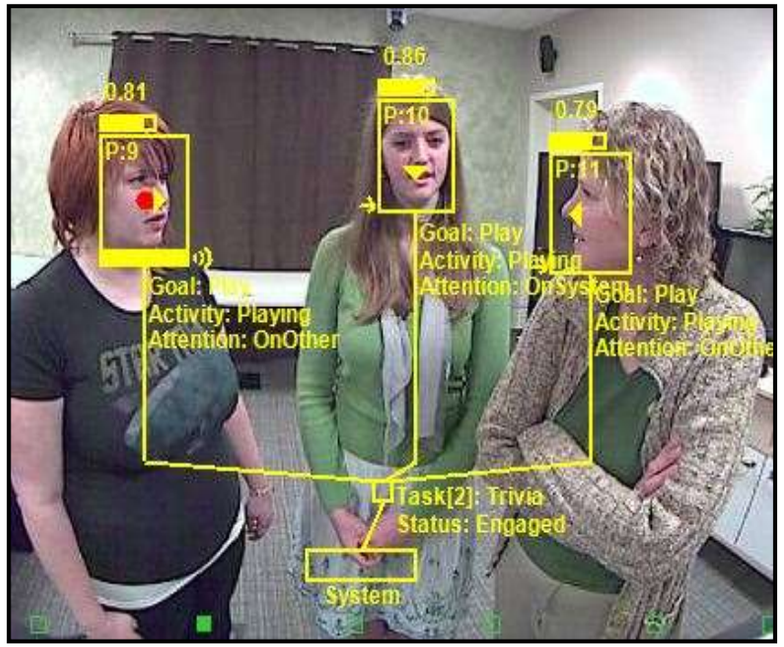
Platform for Situated Intelligence

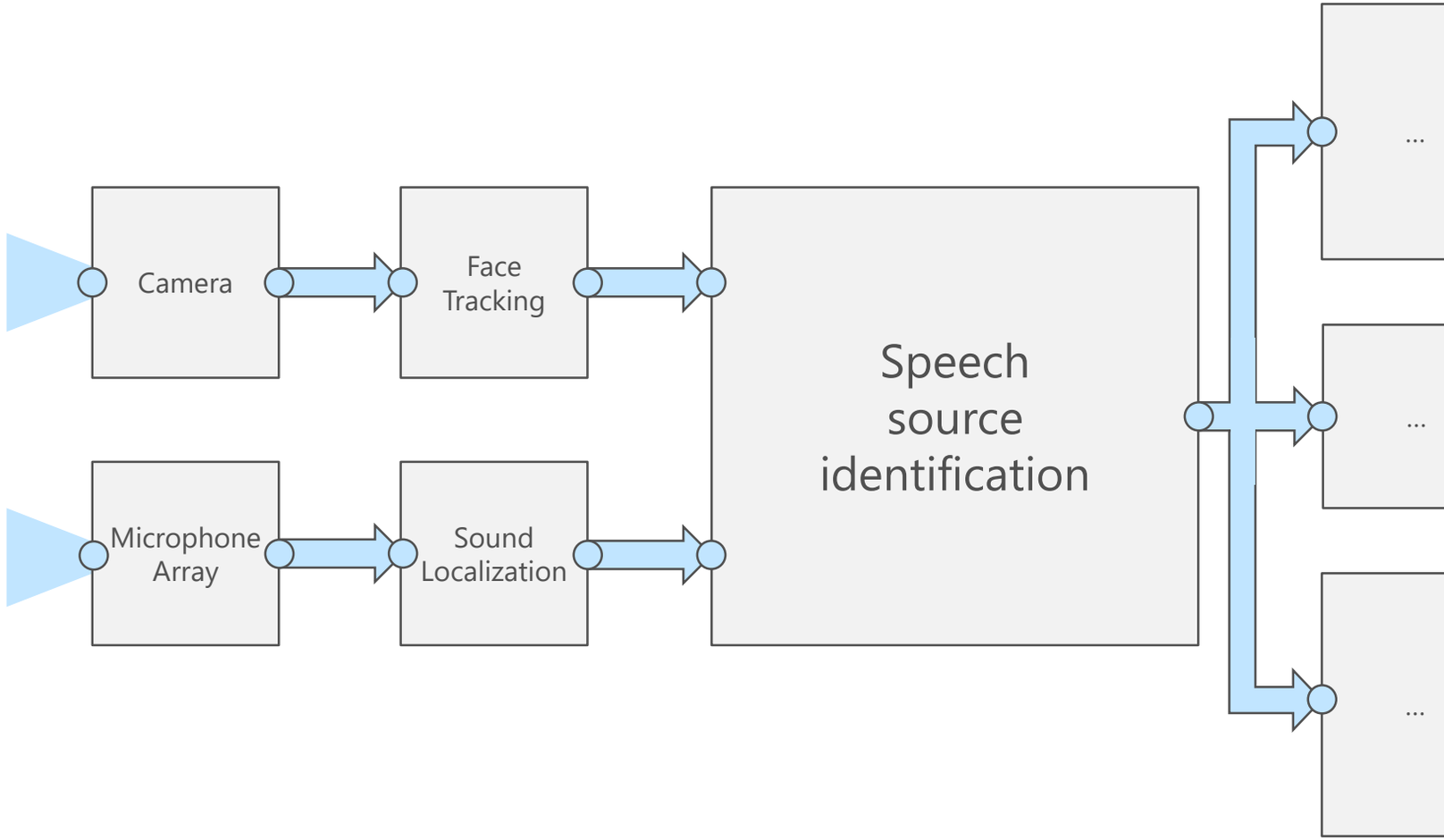


Basic primitives and programming model

Runtime: parallel coordinated computation

Connect components

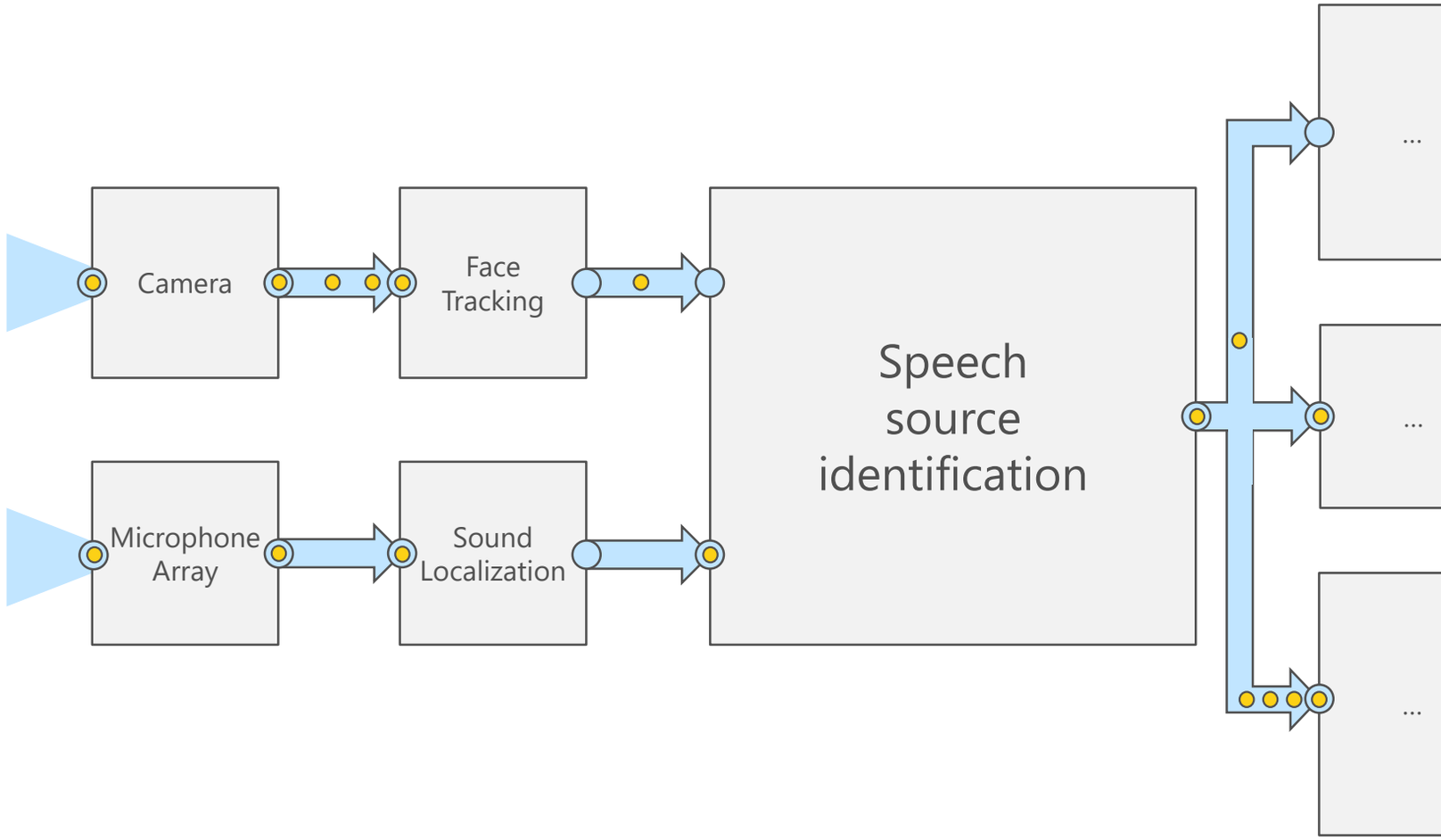




Runtime: parallel coordinated computation

Connect components

Streams & messages

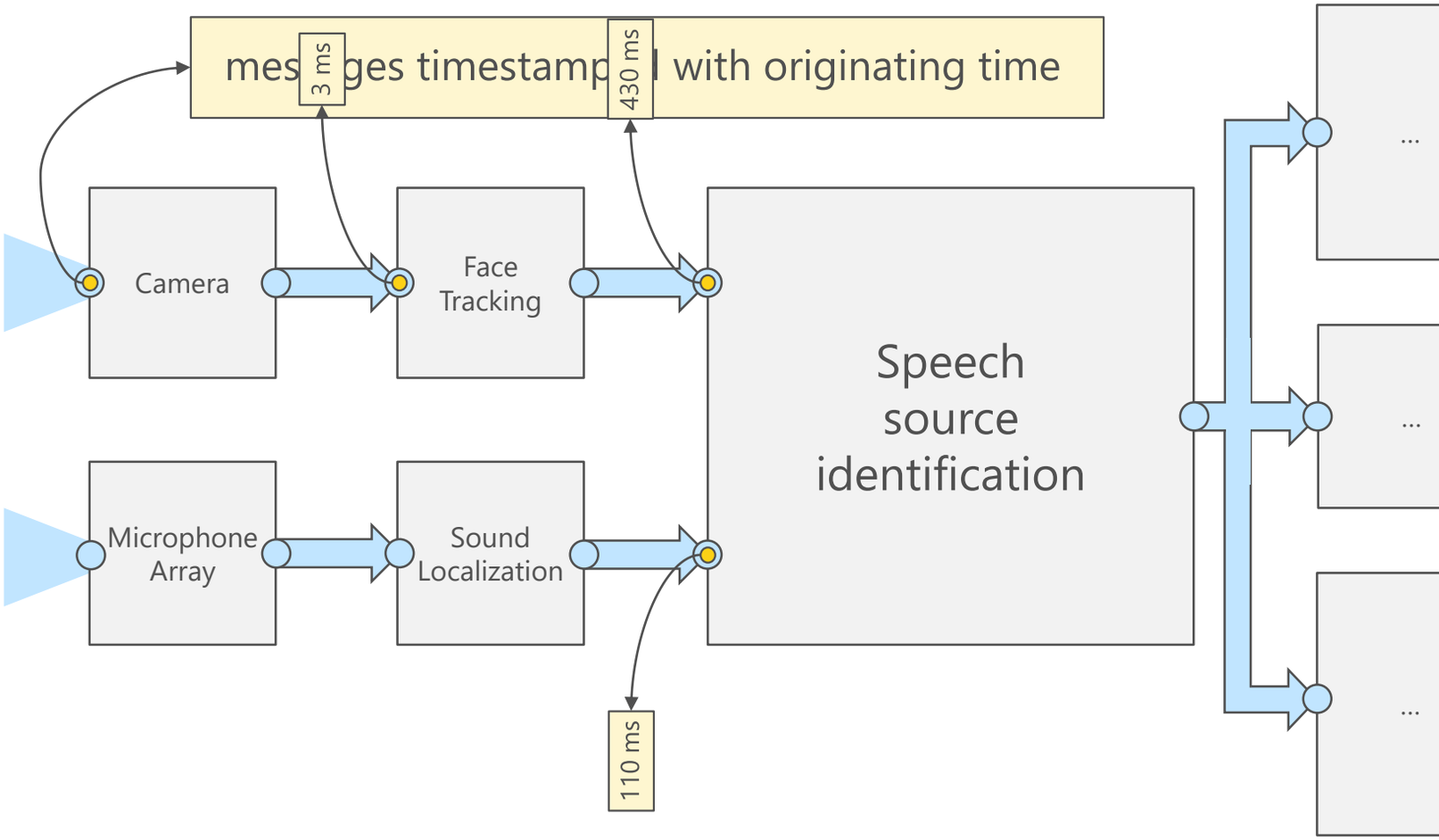


Runtime: parallel coordinated computation

Connect components

Streams & messages

Synchronization



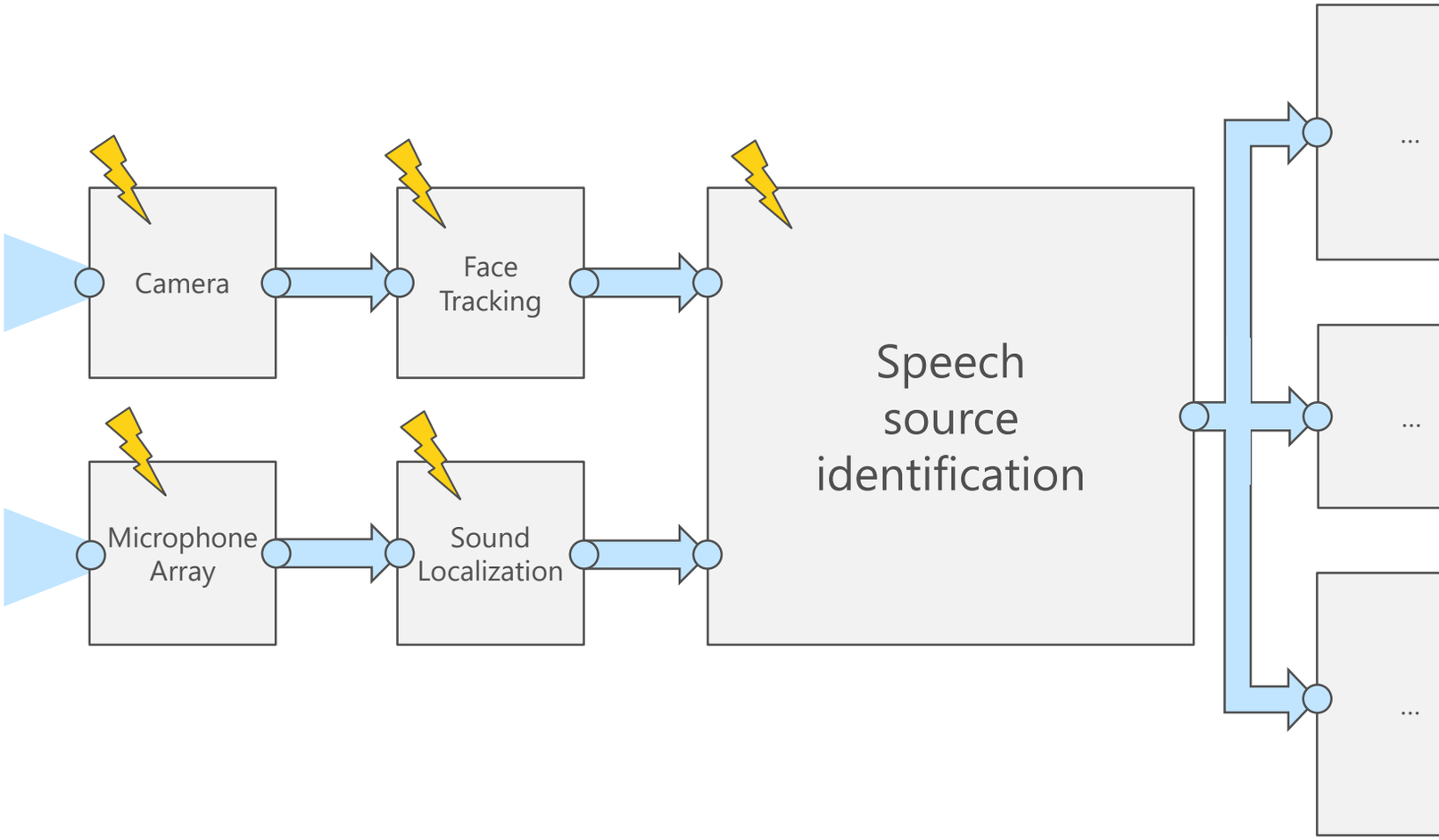
Runtime: parallel coordinated computation

Connect components

Streams & messages

Synchronization

Scheduling



Runtime: parallel coordinated computation

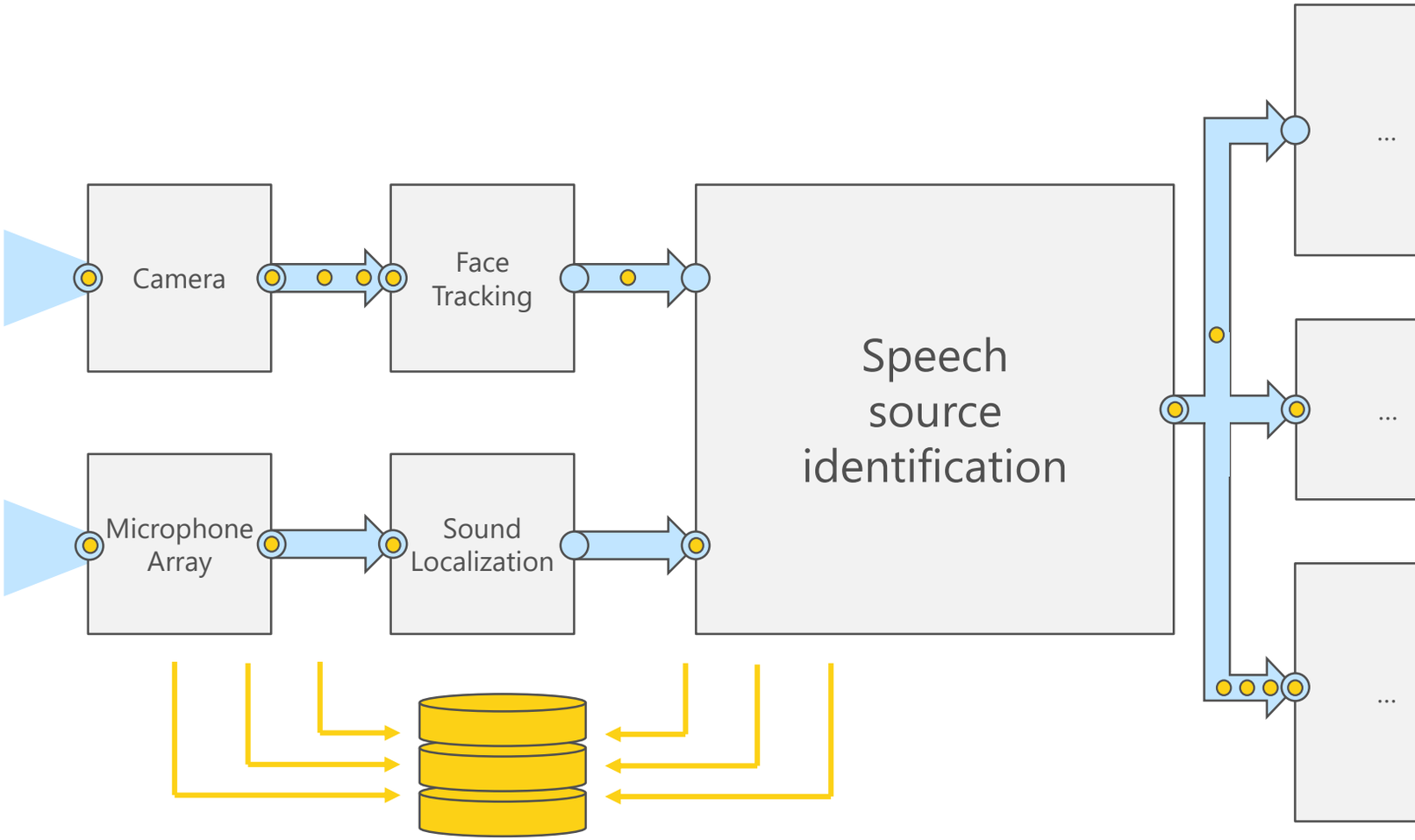
Connect components

Streams & messages

Synchronization

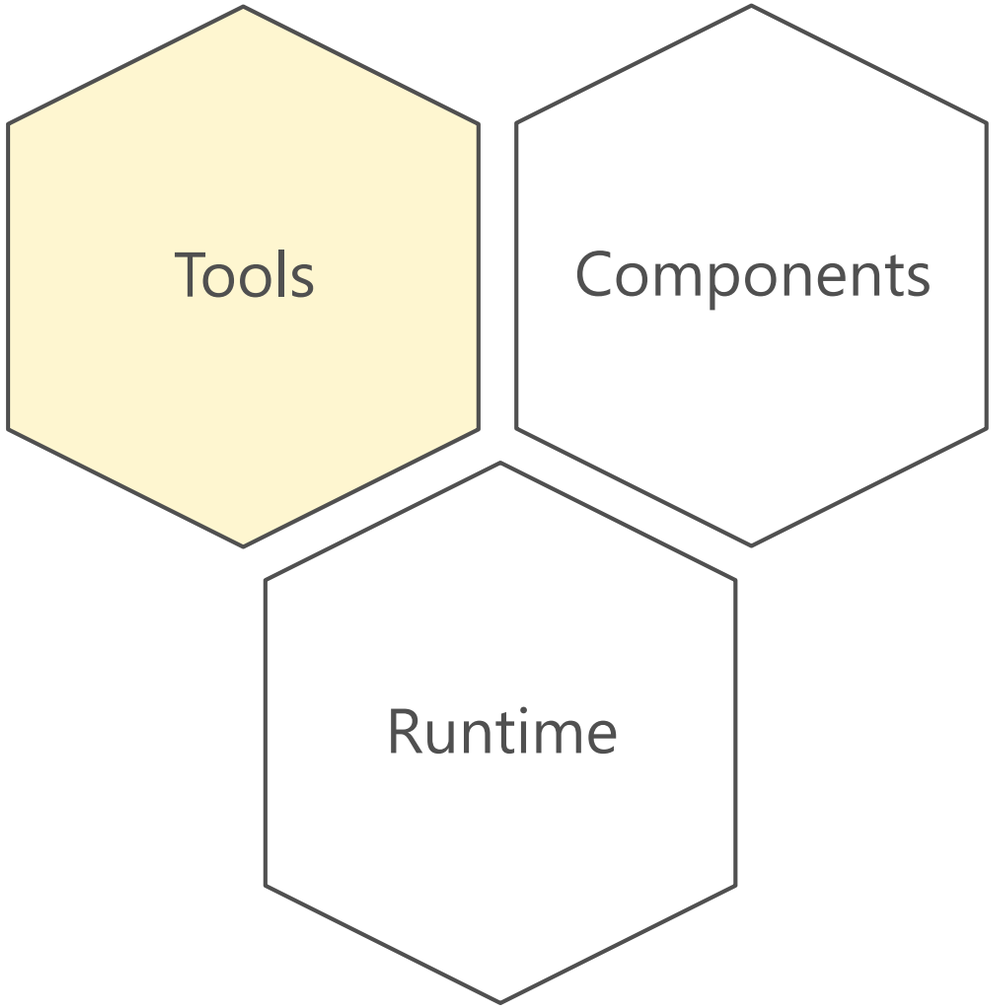
Scheduling

Persistence



Platform for Situated Intelligence

Debugging
Visualization
Machine Learning



Tools

Visualization

Composited visualizations
Navigation
Live mode

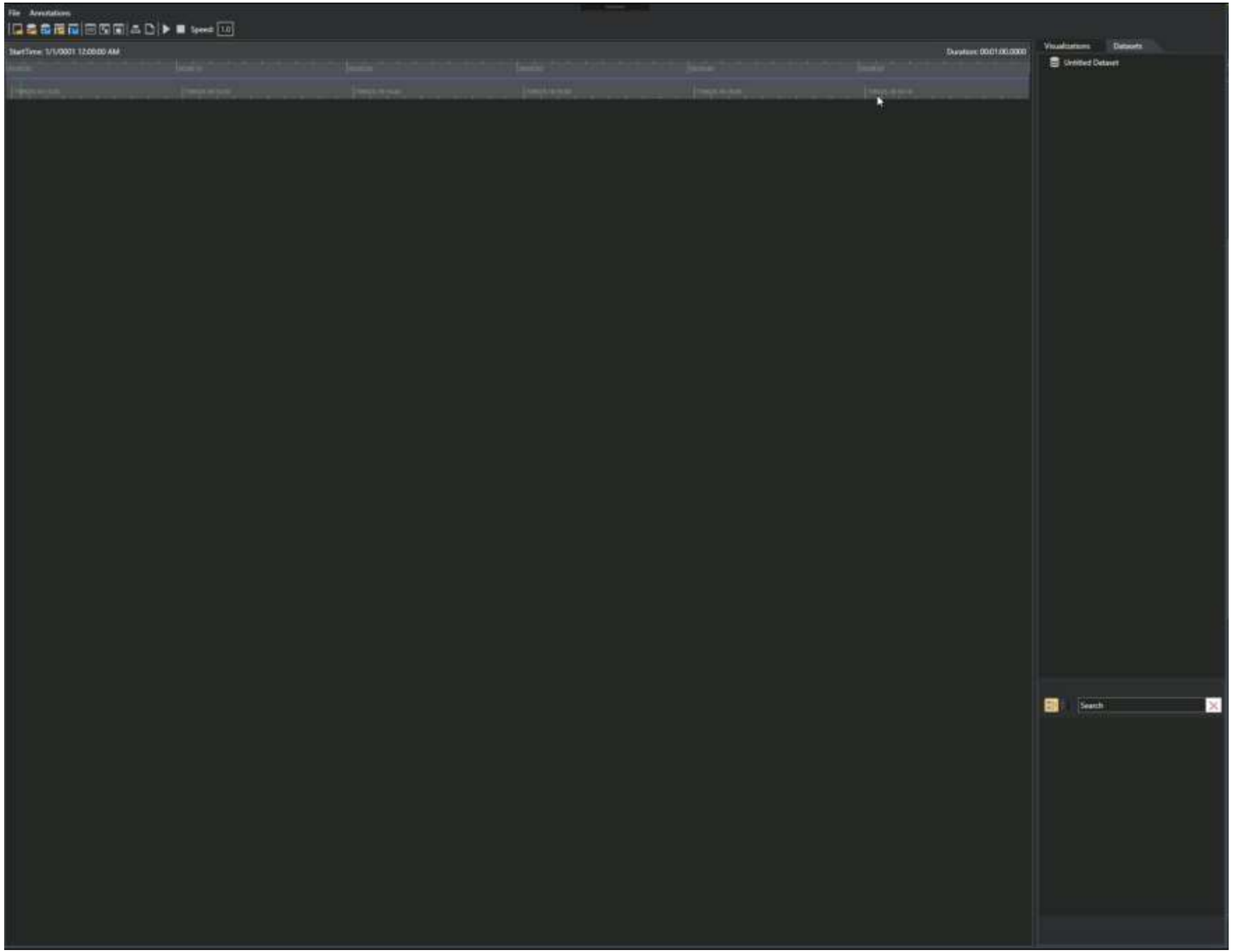
Tools

Visualization

Composited visualizations

Navigation

Live mode

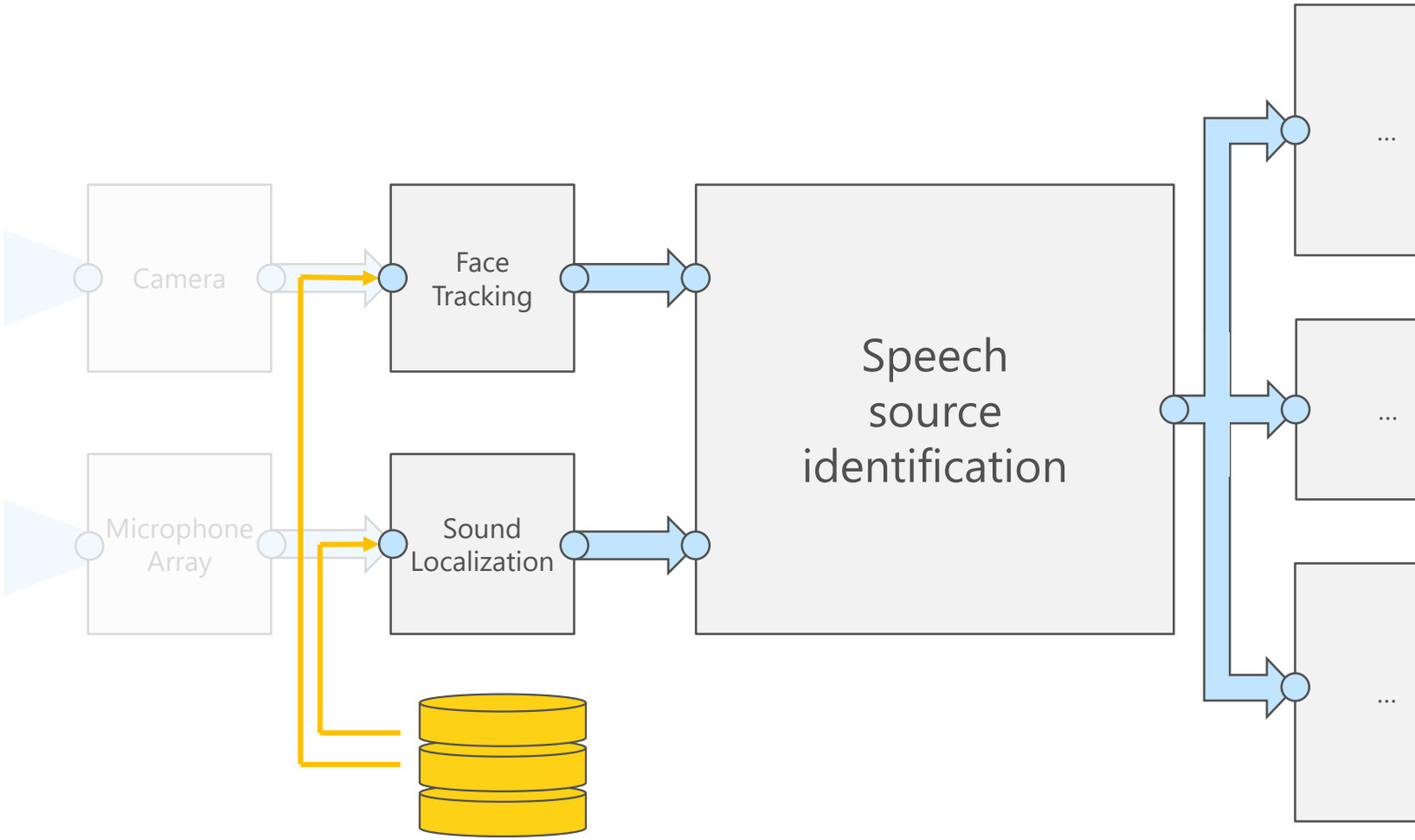


Tools

Visualization

Composited visualizations
Navigation
Live mode

Replay & Experiment



Tools

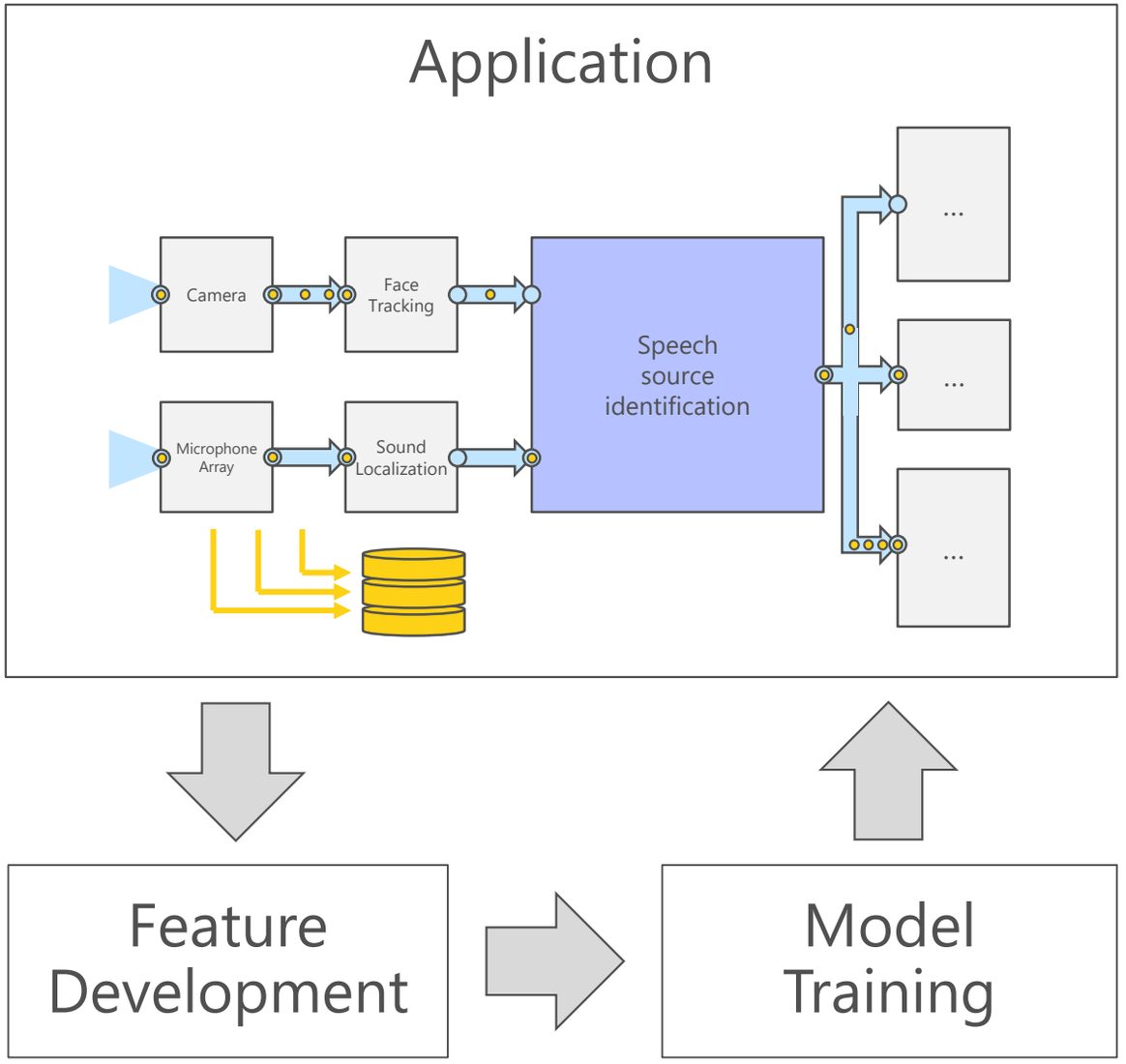
Visualization

Composited visualizations
Navigation
Live mode

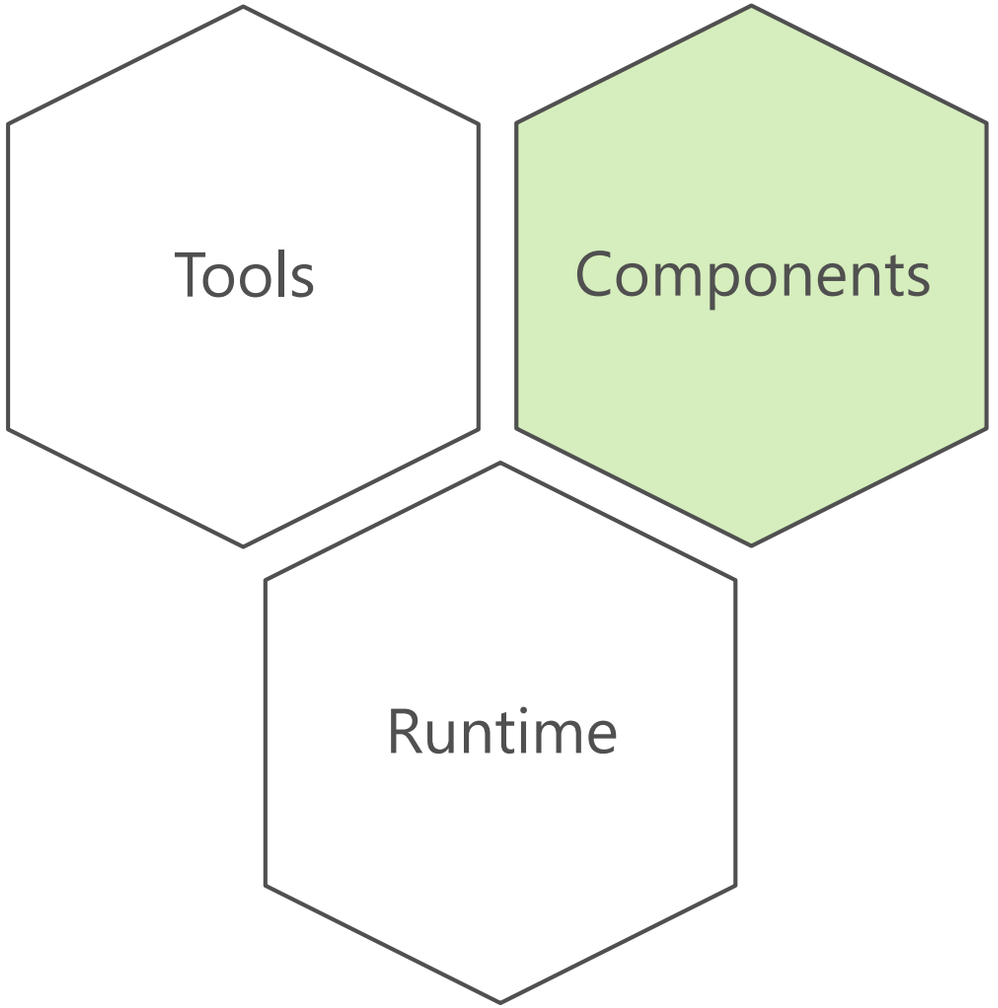
Replay & Experiment

Machine learning

Annotation
Feature construction
End-to-end support



Platform for Situated Intelligence



Repository of
AI technologies

Store for real-time AI components

Wrap existing technologies

Sensors
- Cameras: USB, PointGrey, Kinect
- Microphones: USB, Array
- …

Vision
- Imaging, transforms, optical flow
- Foreground models
- Blob tracking
- …

Speech
- Voice activity detection
- Speech recognition
- Natural language processing
- Natural language generation
- Speech synthesis
- …

…

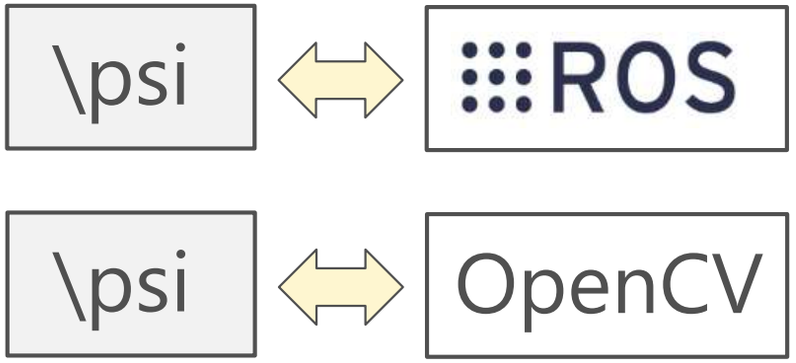
Human perception
- Face tracking & identity
- Gaze tracking
- Body tracking
- Affect recognition
- Attention tracking
- …

Interaction
- Engagement models
- Turn-taking models
- FSM-based controller
- …

Output
- Speech synthesis
- Avatar rendering
- …

…

Bridging & Interop



…

Situated Interaction Toolkit

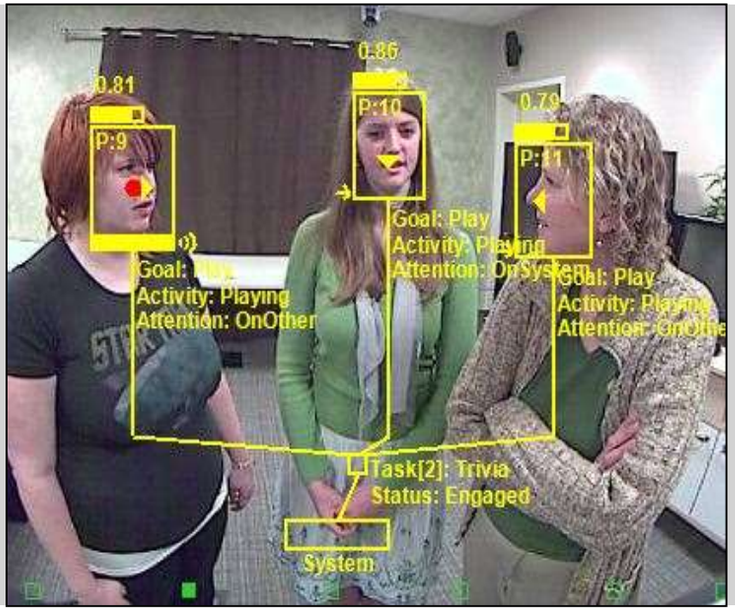
Programming against the “physics” of human activities

identity, attention and interruptability

activities, plans and goals

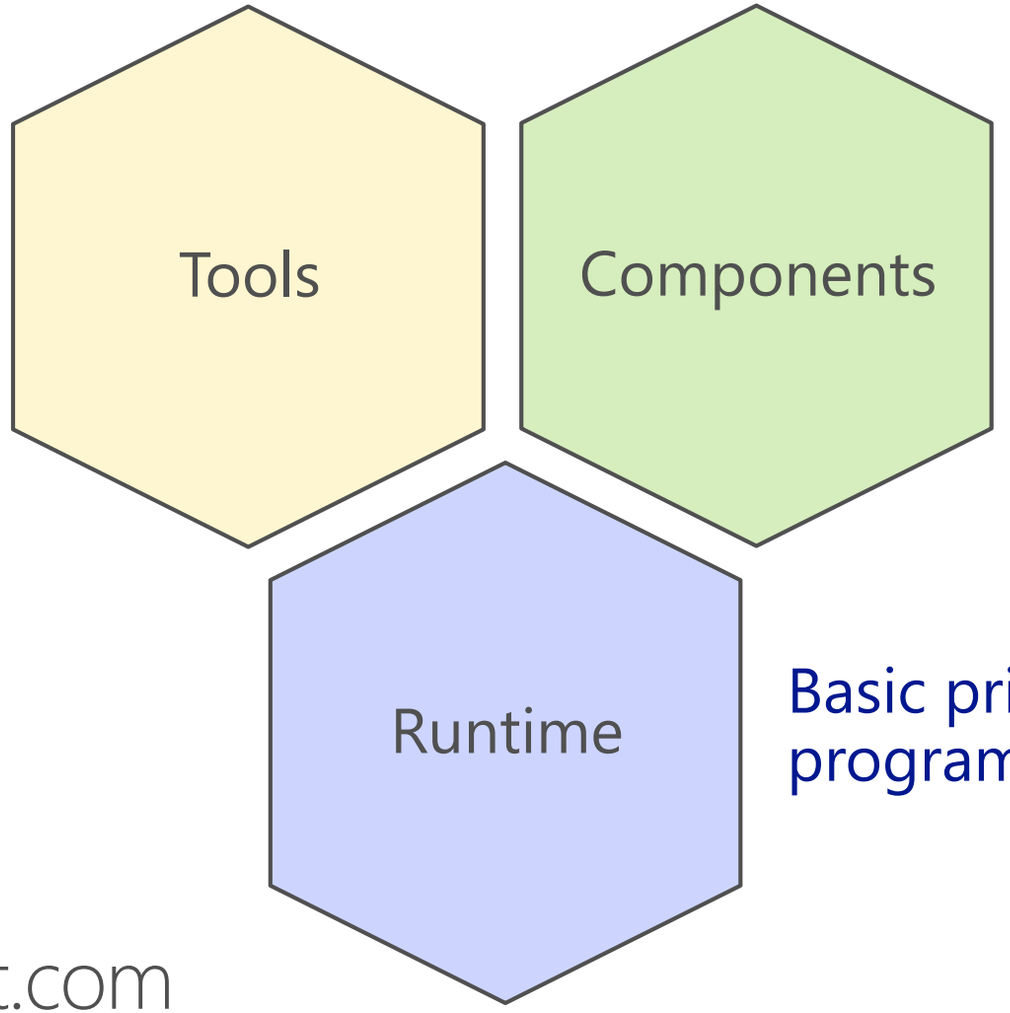
space and time

situated language



Platform for Situated Intelligence

Debugging
Visualization
Machine Learning



Repository of
AI technologies

Basic primitives and
programming model

dbohus@microsoft.com